

Hiding from the midday Sun

Worry about the integrity of the ozone layer — a long-term problem — might vanish if the greenhouse cost of ozone-unfriendly chemicals were charged out at the market rate.

FIRST reports of this season's observations in Europe of atmospheric ozone over the northern part of the Northern Hemisphere (see *Nature* 356, 550; 1992) have been variously interpreted in the general press as reassuring and as a sign that prudent people will henceforth stay indoors, away from the Sun. Not surprisingly, the truth is neither of these things. It is merely that the European Arctic Stratospheric Ozone Experiment (supported by the European Commission and member states) has put out preliminary data suggesting a substantial early-spring-time decline of ozone above several stations in northern latitudes. At Uccle in Belgium, for example, the vertical column held 18 per cent less ozone this January than the January average of recent years. Is that a warning to northerners to stay out of the Sun in the summer that lies ahead?

Most trivially, of course, the warning is too late; ozone seems to have been unusually depleted over northern latitudes in December, January and to a lesser extent in February. By last month (March), ozone concentrations were more or less back (within the inter-annual variation) to normal. There is as yet no reason to fear that sunbathing in the summer months to come will be dramatically more dangerous than in recent years. Indeed, even unplanned exposure to the Sun in the months now past is unlikely to have been much of an extra hazard; it is good luck of a kind that in the late-winter months of exceptional ozone depletion, the insulation (radiation flux per unit area) is reduced and the path-length of solar radiation through the ultraviolet-cleansing atmosphere is increased.

The European ozone project has also been at pains to catalogue the reasons why the northern winter past may have been exceptional. The eruption of Mount Pinatubo has been a dramatic influence on the stratosphere, increasing the concentration of aerosol in the lower stratosphere by a factor of ten. As a consequence or for some other reason, the Arctic concentration of nitrogen oxides (some of which inhibit the catalytic conversion of ozone into oxygen) was anomalously low. And there was exceptional weather in the troposphere below. The northern winter of 1991–92 was probably an exception, but only time will tell.

Just what time will tell is another matter. Seventy years have passed since G.M.B. Dobson built the first instrument for the measurement of atmospheric ozone (by comparing solar absorption at two ultraviolet

frequencies on either side of the high-frequency ozone absorption band), and rather less than that since he showed that most atmospheric ozone is in the stratosphere. Understanding the photochemical reactions that account for the formation of ozone (from oxygen, by ultraviolet radiation), the protection which it affords species like people, and its destruction (by chemicals such as hydroxyl ions, but also by halogen and ClO ions) has required nothing less than an understanding of the chemistry of the atmosphere.

Those who ask whether it is magic that the emergence of problems in physics is so often just matched by the earlier proof of suitable theorems in mathematics might also wonder why, at least so far, practical problems such as that of ozone are so often narrowly preceded by the required understanding. Whatever the case, ozone and its depletion is now a well-known problem, if one not fully understood. Stable chemicals called chlorofluorocarbons (CFCs), widely used as refrigerants, are photolysed in the stratosphere, yielding chlorine and ClO ions. Ozone should be (and is seen to be) reduced. But extrapolations forward are uncertain. At high latitudes, ozone concentrations may be declining by 8 per cent a decade, but elsewhere, where people prefer to live, the decline may be less rapid. The only certainty is that there is a decline and not the opposite. That is why this journal has always applauded the Montreal Protocol of the Vienna Convention on the protection of the ozone layer; regulating the emission of CFCs is necessary because, while the danger to people and other forms of life is still uncertain, the biology scandalously so, nobody can deny the danger.

But will not all this be sorted out at the "Earth Summit" at Rio a few weeks from now? Unfortunately, that seems not on the cards. The next meeting of the signatories of the Montreal agreement is in November, and will consist mostly of the persuasion of the recalcitrant to follow others in banning the production of CFCs. (The non-signatories, who will not be represented, will not have to listen.)

That is why the issue should be brought up at Rio, but in another guise. CFCs are also greenhouse gases and, as such, many of them may be 100,000 times more effective, weight for weight, than carbon dioxide. But organizations such as the European Community are talking of a "carbon tax" on fossil fuels of \$1 a gallon or therea-



bouts. The logic of the numbers is that it would be just as worthwhile to spend \$100,000 to keep a kilogram of CFCs out of the atmosphere. If only the gathering at Rio would offer a bounty of that order for the recovery of CFCs, it would kill three birds with one stone: governments would have an incentive to sign the Montreal Protocol, the greenhouse problem would be abated and the surface of the Earth would be quickly stripped clean of discarded refrigerators. Might not that prospect allow distinguished delegates (in UN-speak) stretch out on Rio's beaches with a sense of a job well done? □

Thirty years back

Britain's new government has an unexpectedly intelligent recipe for managing research.

NOT for the first time in Britain, a newly elected government has stolen its defeated opponent's clothes. Last week, the new Major government quietly launched a radical reorganization of its mechanism for the support of science, and for basic science in particular. Despite masterly pretence, before and during the general election, that all is well, Major has made the organizational change first demanded in 1981 by the House of Lords Select Committee on Science and Technology, and widely echoed (by *Nature's* manifesto, for example) ever since. Indeed, there will be a minister (Mr William Waldegrave) part-time in charge of research spending (he also has to worry about redress for short-changed consumers of public services). He will be sited in the Cabinet Office, which implies access to the prime minister, as the defeated Labour Party had been promising. Labour's complaints of robbery, if any, will quickly be drowned by huzzas for consensuality.

This change takes the politics of British science back 30 years, when there was last a minister of science (part-time), even down to the details. (Waldegrave, like his predecessor, Lord Hailsham, is also a fellow of All Souls' College, Oxford.) It remains to be seen whether the new arrangements will include a more effective and independent advisory body than the present Advisory Board for the Research Councils, whose chief task (dividing the annual budget-pie among five research councils and the Royal Society) is not enough to keep able people constructively occupied without bickering. The need is for a more independent council, free to pronounce on what it chooses, preferably in public. Lord Todd, chairman from 1952 to 1964 of what was then the Advisory Council on Science Policy (ACSP), would no doubt happily advise.

There are several good reasons why Waldegrave should give that wheeze a try, not the least of which is that the present pitiful morale of the British research enterprise owes much to the way in which its professionals have been pushed about, in the past decade, by seemingly arbitrary decisions by ministers and their acolytes. Throughout that time, there has been no sense of dialogue with the

government, but a sense only that orders have been issued and must be obeyed. Nothing could help more in present circumstances than giving a dozen or so people a statutory right to bring up with the minister questions that seem important to the research community. But the dialogue would, in the nature of these things, prove fruitful (as the experience of the ACSP showed). A well-run council would not just be a lightning-rod for discontent, but a way of telling more accurately than for a decade what should be done. □

Leave the man alone

The argument between researchers and the animal rights movements go back a long way.

"I KNOW that physiology cannot progress except by means of experiments on living animals, and I feel the deepest conviction that he who retards the progress of physiology commits a crime against mankind." Who could possibly have written that? There are a few intrinsic clues. Not merely is the message not politically correct, but neither is the language: the author's use of "he who" rather than "anyone who" is archaic. The use of "physiology" suggests a date before "pharmacology" was invented. And the liberal use of "progress", both as a verb and as a noun, would place the text (which is in English) somewhere between 1832 (the Great Reform Bill) and 1918 (the end of the Age of Innocence). But who said it?

Those who have read this far will recognize that they have been led up the garden path. Charles Darwin was the author. Professor Paul J. Whalen from the University of Vermont at Burlington has written to draw attention to the appearance of this opinion in *Nature* on 21 April exactly a century ago (45, 583; 1892). Whalen's worry is that Darwin is now often referred to by those active in the cause of animal rights as one who would naturally have been a paid-up member of their organizations if he had not inconsiderately died: for was it not Darwin's triumph to have shown that the difference between people and animals is one only of degree?

We forget that, while time passes, some contentious issues remain unchanged. Darwin, in his letter, refers to the "agitation against physiologists" in earlier years and says, "I have all my life been a strong advocate for humanity to animals, and have done what I could in my writings to enforce this duty." He wonders "how many lives and what a fearful amount of suffering have been saved by the knowledge of parasitic worms through the experiments of Virchow and others on living animals". That was a century ago. It may be too much to ask that the extremists should spend the time between their demonstrations and laboratory burglaries in totting up the numbers of lives since saved, and the suffering avoided (now quantifiable in dollars through damage suits), but will they not please plead Darwin's support for their disruptive causes? □

US prepares to adopt world patent standards

- First-to-invent rule would be dropped
- Talks, bills seek harmony with Europe, Japan

Washington & Tokyo

LAST year, as the US Patent Office threw a celebration to mark the issuance of its five millionth patent, it neglected to mention one embarrassing detail. Because of a fundamental incompatibility between the patent laws of the United States and the rest of the world, the University of Florida researchers who had filed for the patent had inadvertently lost all their non-US rights. They had published their invention — a way to turn biomass into ethanol — before filing for a patent. That is fine under US law, but it voids their claim anywhere else in the world.

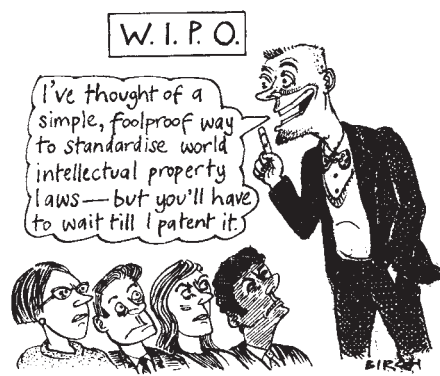
Unfortunately for US researchers, such setbacks are only too common. But an end may be in sight. Pressured by pharmaceutical and biotechnology companies, among others, the United States is preparing to harmonize its system with those of Europe and Japan.

The proposed changes — contained in both ongoing international negotiations and bills pending before Congress — are intended to preserve the free flow of information by retaining those aspects of the US system that allow scientists to publish their results before they are ready to file a patent. But in making concessions to conform to world standard, US officials also expect Europe and Japan to change their systems to give US inventors a better shot at foreign markets.

US patent law is based on the principle of 'first to invent'. No matter when a patent application is filed or who files it, US property rights go to the person who can prove (with a notarized laboratory

notebook or the like) that they made the invention first. A one-year grace period gives US inventors the opportunity to publish first and file later. But in Europe, Japan, Canada — everywhere else in the world except for the Philippines, in fact — the rule is 'first to file'.

As a result, hundreds of US inventors each year lose all rights outside of the United States by publishing accounts of



their discoveries in scientific journals. Even variations on the invention are often lost when foreign companies see the articles and beat the inventors to the European or Japanese patent offices with modified versions of the published work.

Negotiations now underway would change all that. Chief among them is replacing the first-to-invent policy with a first-to-file. In exchange for this concession, the United States wants Europe and Japan to adopt a one-year grace period and policies that would shorten patent reviews

and close loopholes. Under these rules, inventors could still publish (and file for a patent within a year), or file before publishing. But they would no longer be able to use a notarized laboratory notebook to contest a patent filed by somebody else.

Although the spirit of reform is strong, working out an agreement on the details is not easy. International patent negotiations are taking place on parallel tracks: one is under the auspices of the United Nations' World Intellectual Property Organization (WIPO), and the second is part of the continuing General Agreements on Tariffs and Trade (GATT) negotiations. But the GATT talks are deadlocked, and the WIPO negotiations are tangled in the concerns of the developing world.

Nevertheless, momentum for change is building within the United States. Earlier this month, bills to harmonize US patent law with foreign systems were introduced by Senator Dennis DeConcini (Democrat, Arizona), the chairman of the patents, copyrights, and trademarks subcommittee of the Senate Judiciary Committee, and Representative William Hughes (Democrat, New Jersey), chairman of the intellectual property and judicial administration subcommittee of the House Judiciary Committee. Next week (30 April) the two legislators will hold a joint hearing on the subject.

The bills offer several changes to current US law, many of which the US Patent Office has already proposed. Overall, they aim to bring US patent law in line with a future international treaty. If passed, the bills would modify the US law to include:

- First-to-file: as a practical matter, most large US companies already file patent applications in Europe, Japan and the United States before publication to avoid losing their foreign patents rights. But smaller companies and universities often cannot afford such expensive measures (a US patent application typically costs

(continued on page 646)

Patent office drops plan to raise fees

AFTER failing twice to convince Congress that small-scale inventors do not deserve a price break, the US Patent and Trademark Office (PTO) has dropped its opposition to such a discount.

Harry Manbeck, who is stepping down next week as patents commissioner, says that his office will not request a large increase in the fees paid by small-scale inventors over the 17-year life of their patents in legislation that is being drafted to reauthorize the agency. In 1990 and 1991 the patent office tried to equalize the fees paid by all inventors, which are currently twice as high for big companies. Instead, Manbeck says, the agency will ask only for a modest increase, tied to the cost-of-living index, in a bill expected to be introduced next month.

The PTO's change of heart is a victory for individuals, universities, non-profit research organizations and small businesses. Such entities now pay on average a total of \$3,600 to retain a patent over its lifetime. Inventors who fail to pay maintenance fees, due during the fourth, eighth and twelfth

years forfeit their patent.

The patent office had argued that the discount was needed to help it become self-supporting, as decreed by the 1990 budget agreement between the Bush administration and Congress that barred any federal subsidies. Patent officials said that the discount, in effect, forced large companies to subsidize the patent rights of small-scale inventors.

But small inventors convinced Congress that a higher maintenance fees would weaken the US economy in the long run by forcing them to abandon promising discoveries. Even large companies defended the discount, perhaps in recognition of all the good ideas that have originated in a university laboratory or private backyard.

The PTO has learned its lesson, says Manbeck. "Now that it's been considered in Congress", he says, "the Administration is perfectly satisfied to continue the current situation".

Traci Watson

(continued from page 645)
between \$7,000 and \$10,000).

■ **Grace period:** to encourage rapid publication of scientific results, the bills would offer a one-year grace period after publication during which an inventor can file a patent application. Japan currently has a six-month grace period; Europe has none. The bills would also institute a low-cost 'place-holder' process that would allow inventors to file a bare-bones application for initial review.

■ **Publication after 18 months:** US law allows a patent to be kept secret until it is issued. In the case of a patent on the integrated circuit, the inventor used various legal devices to keep it out of the hands of the computer industry for 16 years. Japan and Europe want the US Patent Office to make a patent application publicly available after 18 months (as the Japanese system requires) to avoid such situations.

■ **Breadth of claims:** Japan, in particular, has traditionally allowed inventors to claim only a very narrow patent — one variation of a chemical structure, for example. This has forced foreign companies to cross-license technologies to Japanese companies. Recently, however, newly industrialized countries such as Korea and Taiwan have begun taking advantage of the narrow claims to move in on Japanese inventions. Japan now looks more favourably on permitting broader claims — known as a 'doctrine of equivalents' — as part of a harmonization treaty.

Although the proposed legislation reflects much of the current thinking at the US Patent Office and among US industry, its real purpose is to build a consensus for the international negotiations. Delegates at the WIPO talks have also discussed a 'native-tongue' provision that would incorporate both the original and the translated versions of an invention (to avoid a situation that often comes up in Japan, in which a faulty translation can invalidate an entire patent). And US companies are seeking a change in the Japanese law that would make it harder for companies to tie up competing patent application for years by filing multiple 'oppositions' during the evaluation period. Japanese companies are understandably resisting this effort.

WIPO negotiators are hoping to achieve a treaty by the summer of 1993, but the talks are slowed by the one-nation, one-vote rule. That is less of a concern within GATT, in which trade determines the number of votes allotted to each country and where the agenda is set by the industrial world.

Experience has shown that such talks invariably take longer than expected. But thanks to a consensus on the need for greater harmony, the remaining question is not whether, but rather how much, the United States, Europe and Japan are willing to bend to fit the mould.

**Christopher Anderson
& David Swinbanks**

Out of many, one

London

THE Conservative party, fresh from its narrow electoral victory, has turned away from its philosophy of decentralization in a restructuring of science and technology policy in Britain.

Under a new arrangement announced last week, responsibility for science and technology has been transferred from the Department of Education and Science (DES) to the Cabinet Office and the new Office of Science and Technology (OST). The office will be run by the government's chief science advisor, William Stewart, and will incorporate both the Advisory Council on Science and Technology (ACOST) and the Advisory Board to the Research Councils (ABRC). ACOST advises the government on priorities in home and international science, while the ABRC communicates the needs of the research councils to the government and allocates the resulting slice of the budget.

The new body falls under the control of the Chancellor of the Duchy of Lancaster, William Waldegrave, who was previously Secretary of State for Health. This means that for the first time since 1959 there is a minister of Cabinet rank with responsibility for science.

In a statement released last week, Waldegrave indicated that the changes were motivated by the wish to have a high-ranking minister representing British science in Brussels, and reflect the government's continuing interest in science. Although the changes take effect immediately, an agenda will not be set until a reconvened Parliament approves the reorganization.

Waldegrave is also responsible for the civil service and the Citizen's Charter, an on-going programme for setting standards of public service. As such, securing the annual allocation of funds for the research councils will be his only financial mission.

Waldegrave will be assisted in his duties by Robert Jackson, who has moved from the Department of Employment. Jackson had been responsible for higher education and science when it was part of the DES from 1987 to 1990. Although he made some enemies in the scientific community (the pressure group British Scientists Abroad has called him the 'rottweiler of academia' because of his insistence that the 'brain drain' was a myth), most commentators are pleased that the position has gone to an experienced administrator.

The creation of the Office of Science and Technology has been greeted with guarded enthusiasm from the scientific community. The chief criticism has been that the new structure further widens the gulf between the universities and polytechnics and research.

"Having someone in the Cabinet whose only financial responsibility is science is a great step ahead," commented Sir Eric Ash, rector of Imperial College, London. "But one part I do find troublesome is that the Department of Education will retain responsibility for some of the science budget. This is a problem that needs to be sorted out."

Many commentators share the view that keeping open the lines of communication between science and education will be of paramount importance. But they point out that the gap has been widening and that the previous secretary of state for education, Kenneth Clarke, gave short shrift to science.

"With the problems afflicting science at the present time, I am almost glad to see science removed from the DES, although there are dangers in that", commented Lord Flowers, chairman of the select committee on science and technology in the House of Lords. Reports from this committee have criticized the government's handling of science funding.

Lord Flowers also said that the relationships between the research councils and other government departments — between the Agricultural and Food Research Council and the Ministry of Agriculture, or the Medical Research Council and the Department of Health for instance — might also benefit from increased attention. "I believe it needs a lot of thought", he says about those ties. "At the moment, the Prime Minister can decide to do anything he wants. He could prevent a lot of internecine arguments."

Given that science was barely an issue in the recent election, the creation of the OST has come as a surprise to many people in the United Kingdom. Not the least of these is the Labour Party, whose pre-election manifesto proposed something remarkably similar to the OST.

"Imitation is the sincerest form of flattery," commented Jeremy Bray, Labour's science and technology spokesman, in welcoming the move. However, he maintained that the government had adopted the structure without also taking on the party's underlying strategy for science.

"Nothing has been said about the coordination of research and development across departments" says Bray. "Nothing has been said about integrating policy research and statistical services for government into the Office."

Whether the OST is the whole story or is just the first of a number of adjustments to the way that the government handles science and technology will not be clear for at least the next few months. Word from the inside has it that the shaking-down process has some way to go before the results are clear.

Ian Mundell

Political feud blocks funds

A DISPUTE between the Italian Space Agency (ASI) and its advisory committee over how to spend the agency's budget is blocking the flow of tens of millions of dollars to space scientists. The political battle has delayed the analysis of data from orbiting instruments, jeopardized Italy's commitments to its international partners and damaged morale within the space science community.

The impasse began last May, when the science advisory committee to (ASI) put forward a five-year plan. The plan covered collaborations with its neighbours in the European Space Agency (ESA), the United States and the former Soviet Union, as well as a national programme.

ASI, created in 1988, is required by law to spend 15 per cent of its budget on research. For 1991 that was 68 billion lira (US\$55 million). And the Science Committee, which is elected by ASI, is supposed to recommend how that money should be spent. But last autumn that process broke down.

Between May and November, ASI reduced the amount going to science by more than half. It did so in two ways. ASI took out its contribution to ESA before computing the 15-per-cent share for science. It also allocated some 20 billion lira to projects on an X-ray satellite due to be launched next year.

Angered by this erosion of its responsibilities, the Science Committee refused to accept this reduced budget, saying it would damage the long-term plan. Instead, it submitted a proposal for 74 billion lira, a bit higher than its 15-per-cent share, which ASI rejected. ASI is now waiting for a new proposal in line with available funds.

The debate is not just about money. The impasse reflects unresolved questions about the proper relationship between ASI and the Science Committee. Many scientists believe that the only solution is to dissolve the committee and start again.

Committee chairman Remo Ruffini, of the University of Rome, argues that any sizable reduction in projects funded would threaten Italy's future in space research. Although he says he is prepared to reduce his proposal by 20 per cent, he is determined to continue the committee's fight for "the constitutional right to have their full budget returned".

But some committee members feel that Ruffini's hard line threatens Italy's future in space science, and that what is needed is a compromise offer that ASI would accept. Committee member Franco Pacini of the Florence Observatory, who has called for Ruffini's resignation, says that the time to oppose the budget cuts was last summer, when it was announced. Instead, he says, the committee adopted a head-in-

the-sand attitude that damaged its effectiveness. But he also feels that political and organizational problems within ASI have blurred the financial picture unnecessarily.

Criticisms of the Science Committee are widespread. Unlike advisory bodies in other countries, Italy's committee contains a mixture of pure and applied scientists with differing interests. In addition, ASI has been criticized for letting politics influence its selection of members. Scientists also feel that the absence of a formal peer review system and any mechanism for public accountability leaves them with too little representation.

Giovanni Bignami, of the University of Cassino, says that ASI should take immediate steps to make sure that the Science Committee works because it has ultimate responsibility for the projects it funds. Otherwise, he argues, the Committee should be replaced by an advisory body similar to that used by ESA and the US National Aeronautics and Space Administration (NASA), which have appropriate peer review systems and which publish their decisions.

Bignami, principal investigator for the European Photon Imaging Camera, an instrument on the ESA cornerstone mission XMM, is frustrated by the long delay in funding. Although salaries are independent of ASI funding, the impasse has meant that no money is available for travel and hardware.

The cost to major space science projects is very high. Pier Luigi Bernacca is principal investigator in Italy for the HIPPARCOS project, a satellite that is producing data for a European consortium, the Fundamental Astronomy by Space Technology (FAST). Bernacca says that the backlog of one year's worth of data is causing problems for his consortium partners, and he blames those responsible for holding back funds for "putting their personal views above science".

Giancarlo Noci of the University of Florence, whose contribution to the Solar and Heliospheric Observatory (SOHO) mission was reduced because of the funding difficulties, says that the damage can never be repaired fully. The delay has raised not only the cost of the project but also the risks involved by forcing researchers to abandon many of their test models.

The situation is also harmful to Italy's international reputations. In a field where international collaboration is essential, this is both embarrassing and destructive.

In the meantime, ASI and the Science Committee are meeting regularly in the hope of reaching a settlement — forced or otherwise — by May. At the very least, release of some interim funding is expected.

Alison Abbott

New science institute

Tokyo

JAPAN'S new International Institute for Advanced Studies (IIAS), a sort of Japanese equivalent of the Institute for Advanced Studies in Princeton, New Jersey, continues to break new ground in Japan by attracting large amounts of financial support for basic academic research from private industry and philanthropists.

The latest contribution comes from Shimadzu, a Japanese scientific instrument maker. It has just agreed to donate ¥400 million (US\$3 million) to IIAS over five years in memory of Shosaku Numa, one of Japan's leading molecular biologists who died in February (see *Nature* 356, 193; 1992). The money will be used by a team of three researchers led by Y. Kubo, a former research associate of Numa's, to continue Numa's research on acetylcholine receptors and ion channels, says Saburo Fukui, chairman of the IIAS planning committee.

Japan has few private foundations that support research, and most of them tend to be small and confined by bureaucratic government regulations to narrow fields. But IIAS seems ready to break out of that mould, assuming that the Japanese economy recovers from its recent problems and science is seen as an important national priority.

The institute has so far raised more than ¥5,200 million (about \$40 million) from private contributions, and institute organizers have set the ambitious goal of increasing this to ¥50,000 million by 1994. That goal is tied in part to the institute's planned move next year from temporary offices in Kyoto to the heart of the new Kansai Culture and Science City.

Much of the money already pledged has come from private individuals in the form of cash, real estate and stocks. For example, the late Kazuma Tateishi, founder of Omron Corporation, a manufacturer of control system components, donated one million personal shares in his company to support construction of the new IIAS complex.

The Kansai science 'city' consists at present of only a handful of isolated institutes set in mountains between Kyoto, Nara and Osaka. But the city is expected to have a population of 380,000 people when completed next century after an estimated investment of ¥20,000-¥25,000 million.

IIAS is intended to be the 'brain' of the city. It will have a changing population of about 40 senior academics drawn from all over the world, plus supporting administrative staff. Although research in the life sciences will be a central focus, Fukui says, the institute will also support work in philosophy and mathematics. Topics will be chosen by its faculty as well as by outside contributors from both industry and the government. David Swinbanks

Publish in English, or perish?

Paris

A FRENCH scientist claims that he was denied promotion within the country's largest research organization because none of his work had been published in English. His accusation has reopened a debate between the country's research establishment, which maintains that English is the new international language, and cultural and political forces that are trying to prevent further erosion of the status of their mother tongue.

The unlikely figure at the centre of this debate, which extends into the presidential palace, is Claude Roux, a lichenologist at the Mediterranean Institute of Ecology and Palaeoecology at the Faculty of Science and Technology at Saint-Jerome, Marseille. For the past two years, Roux has applied for promotion to the level of director of research at the Centre National de la Recherche Scientifique (CNRS). He was one of 40-50 scientists who had applied each year for eight such positions.

After his second unsuccessful attempt, Roux says that a member of the jury, Jacques Balandreau of the Institut de l'information scientifique et technique (INIST) in Nancy, told him that "your dossier is very good, but you have no chance of being retained because you have not published in English". The confidential report on Roux's application makes no reference to this as a reason for his unfavourable recommendation.

Roux claims his case is part of a pattern by the CNRS to pressure its scientists into publishing in English. "The CNRS message is very clear — you have no future unless you publish in English." In February, Roux says, the president of the CNRS Commission visited his laboratory and "recommended that we associate with an Anglophone author, for the purposes of publication, even if his role was essentially limited to that of a translator."

The Haut Conseil de la Francophonie, a committee under the patronage of French President François Mitterand, has taken up Roux's cause. In its reply to a letter from Roux, the council does not claim to know the true reason for the CNRS decision but nevertheless warns against "cultural alienation", adding that "we have without doubt no need to use English to contribute to the advancement of science."

In 1989, there was a national outcry when the Pasteur Institute decided to publish its journals in English. It was seen as a belated reaction to a trend within the scientific community to converse in English. This is particularly the case in such fast growing fields as molecular biology, but less so in the area in which Roux works.

The president of the jury that evaluated Roux, Robert Barbault of the University of Paris VI, believes that the CNRS posi-

tion on the importance of English makes perfect sense. "Researchers are not only there to work, but also to diffuse their discoveries", he says. "This diffusion is most efficient if it is via widely read journals, that is in English. Publishing in English not only promotes French science abroad, but also draws attention to other published works in French. To advise a candidate for the post of director of research to make an effort to publish in English is the logical consequence of this."

Leaving aside Roux's case, some French scientists say that the government needs to enforce a clear, consistent policy on the subject of using French in their work. In a letter to Mitterand, Marcel Barbero of the Faculty of Sciences at Marseilles states "we can no longer accept that the governing boards of the national research organisations, despite the declarations of the Head of State, the government and the ministries, continue to ignore French as an international language of communication and expression in science." He says that the CNRS no longer subsidises scientific journals in French, and that even CNRS conferences held in France are conducted in English.

The CNRS has ignored these directives "because it is more in touch with the scientific realities", says Barbault. "Conferences have to be held in English, and even when we ask the government for cash for translation facilities they don't give it." Roux says such CNRS policies are nonsense, especially when applied to a small field such as lichenology, where "publication in English is not really a consideration".

Whether Roux was qualified for the promotion is, of course, difficult to know. Says Barbault, "not all researchers have

the opportunity to direct a team. The job calls for more than just being a good researcher. You have to be an animator and a leader, too."

A similar ambiguity exists when trying to compare the quality of Roux's French publications with comparable publications in English-language journals. Balandreau says that "I simply told him that, if you want to increase your chances [of promotion], you must publish in English-language journals which are peer-reviewed. Roux has a large number of articles, but they are mostly in provincial publications."

Roux disputes that statement. Of 97 publications, 24 of them in the past three years, he says that 15 have appeared in CNRS class-A peer-reviewed international journals (8 French, 7 foreign); 4 in other foreign peer-reviewed journals with a majority of foreign participation; 41 in French journals with peer-review boards composed mainly of foreign scientists; and 31 in other French peer-reviewed journals. Roux says the CNRS is wrong to place a higher value on non-peer-reviewed journals in English than in peer-reviewed French journals.

Whatever the outcome of the Roux case, the government now seems ready to take action. This month, the high council decided that publication of scientific reviews and educational works in French should be a criterion in the evaluation of researchers. It also announced plans to start a European multilingual peer-reviewed journal along the lines of *Nature* and *Science*. The proposed journal, which is expected to contain a substantial portion of articles in French, has apparently received support from the European Community. Other possibilities include a multilingual *Science Citation Index* and a council office in New York to complement the existing one in Geneva.

David Bakewell

AIDS THERAPIES

Kemron's secret: it doesn't work

Washington

EVER since a Kenyan researcher took the results of an uncontrolled clinical trial to mean that a dilute oral alpha interferon solution (known as Kemron) could halt AIDS, rumours of an 'African AIDS cure' have run rampant. Now the US National Institutes of Health's AIDS Research Advisory Committee (ARAC) has examined preliminary results from 13 independent studies and come to an unsurprising result: Kemron does not work.

In a statement to be released this week, the ARAC says that no independent trials to date have replicated the Kenyan results. The committee strongly recommends that patients now taking Kemron seek alternatives. At its 31 March meeting, the committee examined preliminary data from a World Health Organization review of 13

trials of the drug. The trials failed to confirm the seroconversion and increased CD4 cell counts that the researchers from Kenya Medical Research Institute had originally reported.

None of this, however, is likely to stop the rumours and conspiracy theories. Last month a science adviser for Jerry Brown, the US presidential candidate, said that investigating Kemron would be one of Brown's priorities. And the Gay Men's Health Crisis (GMHC) in New York says it still gets "an amazing" amount of inquiries about the drug.

The African-American press has alleged that negative reports from clinical trials are a plot to suppress the drug, and Kemron is still sold on the US black market and in Kenya.

Christopher Anderson

Restoring shine on the Big Apple

New York City

HOME to seven medical schools and to Rockefeller University, the New York metropolitan area has long been a magnet for biomedical research. But the Big Apple's high cost of living and souring image is making it increasingly hard for that magnet to attract and retain talent.

Last autumn, the deans of the city's medical schools joined with assorted city and state officials and with captains of real estate, finance and the pharmaceutical industry to form the Commission on Biomedical Research and Development. Working under the auspices of the New York Academy of Medicine, the Commission is supposed to find ways to heighten the attraction.

One local group that is already working on the problem is the Aaron Diamond Foundation, a major supporter of medical research in the city. The Diamond Foundation is spending \$16.6 million on a new postdoctoral research fellowship program. Two months ago, the foundation named its first group of 25 fellows who will work for up to three years in the fields of AIDS or drug addiction research with a mentor at one of New York's research institutions.

The fellowships offer an attractive financial package: an annual stipend beginning at \$36,000, an annual research allowance of \$25,000 and an institutional allowance of \$5,000. Yet the number of applications it received was fewer than expected, and there were few takers outside the metropolis.

Just 73 researchers completed the application process, 61 of whom were already working in New York. Of 25 awards, only two went to researchers working outside of the city.

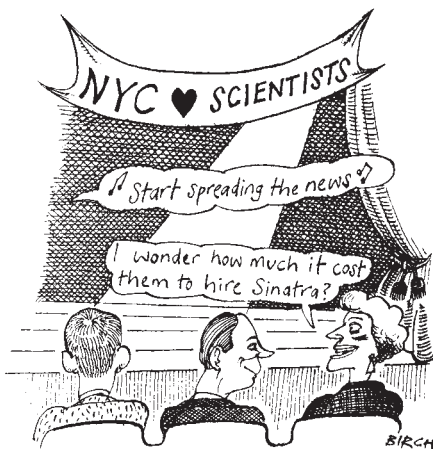
The foundation is modifying its criteria to attract more applicants from outside New York. Researchers will henceforth be required to request work with a mentor or institution with whom they have not worked during the past three years, or to explain why they need further training where they are.

Even so, the program must combat the fact that, for many people, living in New York is not an attractive idea. "I would definitely not raise a family in New York City," says Andrew Henderson, a native Californian who is working at Columbia University's College of Physicians and Surgeons under a grant from the US National Institutes of Health (NIH). Henderson cites such disincentives as the cost of housing in Manhattan, high state and city taxes and the strain of living among such urban problems as crime and homelessness. "In California you're in your car, you're sheltered," says Henderson, who rides the sub-

way to Columbia each day.

On the other hand, the fellowships give researchers who want to remain in New York an opportunity to launch their careers. Diamond fellow JoAnn Difede, who expects to receive her doctorate in psychology from the New School for Social Research in the spring, thought about leaving New York because of the high cost of living. Instead, she will study psychoimmunology with a mentor at Cornell University Medical College. "There's a much better chance that I'll stay in New York now", she says.

The Diamond fellowships are intended in part to close a gap in funding left by the



New York City Health Research Council. The council was formed in the early 1960s and given \$8 million a year — a dollar for every New Yorker — to fund as many as 200 young scientists at a time.

But the program fell victim to the city's fiscal crisis in the mid-1970s. Its successor, the New York State Health Research Council, spends just \$1 million a year on fellowships spread throughout the state.

While John A. Rowe, president of Mount Sinai School of Medicine and chairman of the Commission on Biomedical Research and Development, insists that "biomedicine is not in crisis", the state conducts a declining proportion of the nation's overall biomedical research. According to a study by the Greater New York Hospital Foundation, New York's share of NIH funding fell from 8.1 per cent in 1970 to 6.6 per cent in 1987.

Another indication of the problem is found in a report of a few years ago by the New York State Health Research Council. Although the state has 12 medical schools, compared with four in Massachusetts and eight in California, those states beat New York in the competition for NIH post-doc

grants, training grants, fellowships and small business innovation grants.

New York State slipped by another measure during the late 1980s, according to the Institute for Scientific Information. The number of scientific papers published by New York State researchers from 1985 to 1990, in comparison to the previous five-year period, rose by less than half the rate of the nation as a whole. At the same time, the quality of their work remained high: New York researchers recorded a 16 per cent growth rate for citations, compared with an average increase of 10 per cent for all states.

No one questions that New York still has many attractive features. In addition to the myriad cultural opportunities, researchers like the fact that they have a network of colleagues at other teaching hospitals "within a ten-minute ride of here," says William A. Polf, deputy vice-president of Columbia University Health Sciences.

But back in their own labs, researchers tend to find quarters tight. At Mount Sinai, for example, revenue from research grants and contracts has risen in the last three years alone from \$48 million to \$73 million. Yet there has been no significant increase in space for two decades. The school has vacancies for several department chairs, according to Rowe, but cannot fill them until it acquires more space.

Some institutions are expanding. Rockefeller University is completing a 14-story building, while New York University is putting up a new \$163 million biomolecular sciences building and Columbia is erecting a \$28 million facility with assistance from the city and state. Besides housing university labs, Columbia's new building will make space available to new biotechnology companies.

Columbia's would be the first incubator space for biotechnology in the city, which trails Boston, Washington, and San Francisco as a birthplace for such companies. The Commission on Biomedical Research and Development is studying the problem on the theory that the lack of significant commercial activity may be stifling the growth of biomedical research.

The Commission has also asked the Louis Harris polling organization to survey winners of Federal grants nationwide to learn how young researchers make their decisions about where to work. The plan is to use that information to attract more of them to New York.

"There's a need for a wake-up call," says Rodney Nichols, chief executive officer of the New York Academy of Sciences. "New York cannot any longer assume that its number one status is guaranteed for life."

Mimi Bluestone

Kinder, gentler, more useful mathematics

Washington

THE most prestigious graduate mathematics programmes at US universities attract, naturally enough, the brightest undergraduates. But some of the most highly reputed mathematics departments in the country, having taken in the cream of the crop, do little to enrich it. In some cases, a majority of those enrolled never even finish their degrees.

While some defend this failure rate as proof of the intellectual rigor of the subject, a new report from the National Research Council (NRC)* on doctoral education in the mathematical sciences comes to a simpler but heretical conclusion: many traditional doctoral programmes serve their students, and society in general, badly indeed.

The NRC report, issued last week, was based on scrutiny of graduate programmes in mathematical sciences (including statistics and operations research, for example, as well as the traditional areas of pure and applied mathematics) at ten universities. Their names were concealed to encourage candour among faculty members and, especially, students who were interviewed. These programmes were selected for demonstrating unusual success in better completion and placement rates and higher proportions of women and minority students.

The most evident conclusion from this informal survey was the common-sense one: students invariably did better and were happier in programmes where the faculty took pains to encourage and guide them. They stand apart from what Avner

Friedman, from the Institute of Mathematics and its Applications at the University of Minnesota, calls the "traditional European elitist" system, in which graduate students are expected to produce a thesis after several years' effort or else to pass silently from the scene.

In early drafts of the report, recalls Rhonda Hughes of Bryn Mawr College, there were repeated references to such phrases as "nurturing environment" and "positive learning experience". Their appearance was so startling, she says, that most were cut out in the final report to lessen the shock to the mathematical audience. The gratifying aspect of their study, says Hughes, was that the ingredients needed for success are things that do not cost money; unfortunately, she adds, they are also the things that money can't buy.

Perversely, programmes that are successful in the ways measured by the report still may have trouble attracting students. Experience with her own undergraduates at Bryn Mawr has taught Hughes that 'big name' schools hold a strong attraction despite her warnings that the elite universities will not be comfortable with, and may not even offer, the variety of research disciplines that will best suit the individuals' talents and career prospects. Foreign students likewise face pressure to attend the most famous university their parents have heard of.

But there are indications that the new approach also appeals to students. Programmes that have actively sought to distinguish themselves from the traditional

mathematics departments and which have made efforts to recruit new students find that they can fill as many as 80 per cent of their places with domestic students. That is quite an accomplishment at a time when, overall, more than half of the mathematical sciences doctorates in the United States are awarded to foreign students. Women and minorities comprise as much as 30 per cent of such classes.

Although the aim of the survey was to improve the quality of life for graduate students, socially and perhaps politically useful lessons may come of it. Ronald Douglas of the State University of New York at Stony Brook, and chair of the committee that produced the report, sees the potential for a wholesale change in the way that the mathematical sciences are conducted. Through departments that have made extra-academic contacts, business and industry are waking up to the idea that they need people whose speciality is something called mathematics. Most mathematicians now working for US industry, he says, are disguised as engineers and physicists because mathematics itself sounds too removed from the real world.

The change in strategy extends beyond the newer and less celebrated departments. It is beginning to dawn on faculty at even the most prestigious institutions that a PhD completion rate of less than half is a symbol not of quality but a sign that the system is not working. **David Lindley**

* *Educating mathematical scientists: doctoral study and the postdoctoral experience in the United States.* National Research Council, April 1992.

EUROPEAN RESEARCH FUNDING

Maastricht drives R&D to the market

London

IDEAS with clear commercial potential will receive the most attention in the next five-year plan to cover research and development (R&D) by the European Commission (EC). The approach, described last week by Filippo Pandolfi, commissioner for R&D, is a response to negotiations last December at Maastricht, where the EC heads of government pledged to make their countries more competitive in global markets.

Although science *per se* barely got a look-in at the summit — a move to change the voting system that allocates funds was considered briefly and rejected in the dying minutes of the summit — R&D is getting a boost, says the Commission, to "reflect the ambitions of the treaty".

Among other things, the Commission is concerned about the proportion of gross domestic product that its 12 members spend on R&D. Last year that figure was 2.1 per cent, compared to 2.8 per cent by

the United States (which includes military spending) and 3.5 per cent by Japan. It is also worried that R&D linked to industrial activities does not receive enough attention, and it wonders why European companies find it so difficult to convert their R&D activities into inventions and their inventions into profits.

In response, Pandolfi said that the total resources for Community R&D would be increased from 2,400 million ecu (US\$2,950 million) in 1992 to 4,200 million ecu (US\$5,150 million) in 1997. Even so, the rise falls short of the 5,400 million ecu that Pandolfi said he was aiming for in January (see *Nature* 355, 3; 1992). As well as maintaining existing programmes, this money must also pay for the Commission's intention to become more involved in such 'big science' projects as thermonuclear fusion, global change and the human genome.

While the proposals for these projects will be submitted by individual compa-

nies, the idea is that the technologies involved will benefit European industry as a whole. The Commission is especially interested in supporting such areas as micro-electronics and high-performance computing, advanced molecular biology and environment-friendly industrial technologies.

At the same time, the new approach to near-term research sails close to breaking the rules, which say that the Commission can only support research projects that do not distort competition in the market place. The Commission hopes to uphold that rule by building networks into each project to ensure that each company interested in the results has the chance to exploit them.

Although EC research will be increased significantly if these proposals are approved, it remains a very small part of R&D in Europe. At present it represents less than 4 per cent of the total financial resources allocated to R&D by the member states of the Community. **Ian Mundell**

Animal lib and racial bigotry

SIR — David Wolpert¹ equates animal research with racial bigotry. But the rhetoric of animal rights groups, particularly antivivisectionists, is disturbingly similar to the rhetoric of groups that promote racial violence.

At the centre of antivivisectionism is a core of extreme cynicism and paranoia. Antivivisectionists generally hold that all animal research is meaningless or actually harmful. How can they account for the vast amount of animal research? Their explanation is a vast conspiracy of 'science', 'mass media', 'industry' that actually creates 'new diseases'³. Such global conspiracy theories have always appealed to frightened, alienated and uneducated people looking for simple answers. This cynicism is also the root of the destructive nihilism of the terrorist. Burroughs noted: "where nothing is true . . . therefore everything is permitted"³. The ritualization of cynicism and paranoia is the major theme extending from the ancient Assassin cult^{3,4}, to the Nazi party⁵, to the genocidal Khmer Rouge^{4,6}.

Belief in conspiracy alone is not sinister, but although animal rightists start with the laudable goal of making the treatment of animals more humane, they more often succeed only in dehumanizing other people, especially scientists. This may account for the large number of animal rights activists endorsing vandalism or violence⁶, and the attempted assassination of researchers in England⁷.

Antivivisectionism is also a cult of superiority. Its supporters proclaim themselves to be healthier, morally superior and smarter than any group in human history. With such impressive credentials, they could be considered the new master race.

Antivivisectionists are also contemptuous of the sick. They would end most medical research, because the only reason for being sick is having the wrong lifestyle. This blame-the-victim viewpoint exploits the very real prejudice against people with chronic diseases such as AIDS and even people who have recovered from cancer. Apparently the bright utopia envisaged by the animal rights people will not include those with genetic problems, birth defects and chronic or degenerative diseases. Perhaps the death of these sick people would be the final solution.

Who supports animal research? Wolpert implies a financial motive by stating "No [scientists] are as incensed by the animal rights movement as those whose careers it threatens."¹ I would suggest that the most incensed people are those who feel that the lives of their family members are directly threatened by the

animal rights movement. As a scientist, I understand how animal research was essential for the treatment of my mother's congenital kidney disease, my father's neurosurgery, my best friend's bone marrow transplant and treatment of my classmate's type I diabetes. Perhaps if animal rights activists got out more and volunteered to work with the chronically ill, it might broaden their outlook.

GREGG W. SILK

1170 Evans Trail, Apt 201,
Beltsville, Maryland 20705, USA

1. Wolpert, D. *Nature* **352**, 466 (1991).
2. Handbill—Animal Liberation Network, Hunt Valley, Maryland, USA.
3. Burroughs, W. S. *The Cities of the Red Night* (Holt, Rinehart and Winston, 1981).
4. Morrow, L. *Time* **137**, no. 23, 48–53. (1990).
5. Fromm, E. *Escape From Freedom* (Avon Books, 1941).
6. Storr, A. *Human Destructiveness* (Grove Weidenfeld, New York, 1991).
7. Holden, C. *Science* **253**, 264 (1991).
8. Cherfas, J. *Science* **248**, 1485 (1990).

SIR — Your report (*Nature* **353**, 687; 1991) on the settlement reached by Katie McCabe and Alex Pacheco and the People for the Ethical Treatment of Animals (PETA) in the suit brought against McCabe for her article "Beyond Cruelty" in the *Washingtonian* prompts me to comment on the clarifications and corrections of the article that the *Washingtonian* published. As the magazine stated in an earlier press release, McCabe "... was right on the main ideas, which should continue to be written about".

I believe her article still stands as a powerful indictment of the animal rights movement. In fact, two of the corrections seem to indicate that Alex Pacheco had more interest in using the well-known Silver Spring monkeys as political tools than in caring for them. Although McCabe could not prove that Pacheco had staged a particular photograph of a monkey in a restraining device in Taub's laboratory, it is evident that Pacheco took a picture rather than adjusting the head restraint for the monkey so as to relieve its discomfort.

Similarly, although McCabe could not prove that other photographs (not mentioned in your article) showing uncleaned fecal pans were staged, or that Pacheco himself let laboratory conditions deteriorate in Taub's absence, Pacheco damns himself when he claims he did not clean because "it was not his duty to do so". When animal caretakers are absent, any ordinary researcher knows it becomes his duty.

Still unchallenged is McCabe's most serious charge, that "the only incontrovertible act of animal cruelty was perpetrated by PETA. The three PETA members into whose custody police placed the [Silver Spring] monkeys trans-

ported them 2,000 miles to Florida by truck. After their court-ordered return to Taub, elevated white-blood-cell counts indicated severe stress, and a veterinarian described them as 'one of the most withdrawn and depressed' groups of animals he had ever seen."

The bulk of McCabe's article rightly portrayed a nightmarish outlook — if animal rights organizations prevail in winning their stated objectives — where men, women and children afflicted with diseases for which we now have no treatment or cure will have little hope of ever getting well.

That may be the world PETA wants, but it is not the world wanted by researchers and clinicians devoted to helping people live healthier lives. It is hard to believe that anyone is against doing all we can to help the autistic child caught in his or her silent world, to defeat schizophrenia, to find a vaccine for AIDS, to develop new operations for congenital defects or to realize other key steps that will make the lives of many of our citizens healthier and happier.

FREDERICK K. GOODWIN
(Administrator)

Alcohol, Drug Abuse, and
Mental Health Administration,
Rockville,
Maryland 20857, USA

Analytical language

SIR — J. R. Smythies (*Nature* **355**, 490; 1992) is perhaps too readily dismissive of the belief — held, he says, by many philosophers — that linguistic or logical analysis can contribute to the solution of problems relating to mind and perception. Only observation, hypothesis and experiment can do this, he believes, and I would not wish to challenge this.

But surely observation and analysis of the structure of language, that unique and most complex mode of human behaviour, may tell us something about the structure of the brain and how it functions.

The way the brain is apparently constrained to think (and speak), as reflected in the deep structure of human language, may lead to some understanding of the way the brain is (or is not) wired up and programmed.

Such an understanding may say something about how the brain handles perception and even about the relationship of these neural activities to self-consciousness and the mind. This kind of linguistic analysis would not constitute a poetic approach to cognitive neuroscience, though it was not, of course, Wittgenstein's way.

R. M. FRY

PO Box 387,
Bondi Junction, NSW 2202,
Australia

Anomalies of tissue research

SIR — Barbara J. Culliton ("Needed: fetal tissue research" *Nature* 355, 295; 1992) cites an interesting twist to current US policy banning the use of federal funds for research on fetal tissue transplantation. The twist is that there is no ban on the use of federal funds for fetal tissue research *in vitro* or in animals. But there is an even more fascinating contradiction. I and many others have received federal funds for the past 30 years to conduct research on the use of normal human fetal fibroblasts to produce human virus vaccines (*Am. J. Hygiene* 75, 240; 1962). WI-38, and other normal human fetal cell strains have been, and still are, used in the United States to produce virus vaccines against poliomyelitis, rubella and rabies.

It is undoubtedly true that many of the same people who promulgated and support the ban have also been the benefactors of human fetal tissue research. Although they are not direct recipients of fetal tissue transplants, tens of millions of people in the United States had, or still have, the viral products of these cells coursing through their veins. When the congressional bill to lift the ban reaches the desk of George Bush, as it soon should, he and his administration might well be reminded of this fact when he vetoes the bill, as he promises to do.

LEONARD HAYFLICK

University of California,
San Francisco Medical
School and VAMC,
San Francisco,
California 94121, USA

Rethinking Burt

SIR — There are two quite different issues involved in the controversy over the late Sir Cyril Burt (*Nature* 356, 5; 1992). The first is whether Burt was a "senile liar", a "wicked old fraud" and "the biggest scientific scandal since the Piltdown hoax", as he has been variously described in print, and whether he deserves to be pilloried as among the worst of the offenders, for example in *Betrayers of the Truth* (N. J. Broad & N. Wade, 1982), and *False Prophets* (A. Kohn, 1986), for having deliberately fabricated and falsified data to support his conclusions.

In 1980, the British Psychological Society (BPS) officially supported these allegations of academic fraud, about which doubts have now been raised, and the question is whether they were wrong to have done so. There is clearly a *prima facie* case for looking at the evidence again, and as the BPS is evidently un-

willing to do this itself, some other body should undertake the task. That would most appropriately be done by the British Academy, of which Burt had for many years been a distinguished fellow.

The second question is whether Burt's conclusions were in fact correct, assuming that they were not fraudulent. These were that human intelligence is to a large extent genetically heritable, and not much affected by education or other environmental factors.

Burt's data were almost all collected in the 1920s and 1930s, and as he was a pioneer in this field, it is to be expected that some of his methods may have been deficient by modern standards. So, even if most of the original records had not been destroyed after his death in 1971, Burt's conclusions on the inheritance of intelligence and educability could not now be safely accepted without further research, using the best modern methods. This has of course been done, with results that arguably support Burt's position, though that is controversial. It is anyway something that needs to be investigated further, by educational psychologists and others. The BPS itself, however, might now be better advised to avoid taking any corporate stand on the question.

C. B. GOODHART

Gonville & Caius College,
Cambridge CB2 1JA, UK

Political pressure?

SIR — The fact that the reports of the committee of inquiry and of the field expeditions committee were not made available before V. J. Gupta was reinstated at the Panjab University of Chandigarh (see *Nature* 355, 660; 1992) underlines the possibility that these committees came under political or other pressure — a not uncommon occurrence in India.

MADHUSUDAN G. SONI

National Institute of Nutrition,
Jamai Osmania PO,
Hyderabad 500 007,
India

More than survivors

SIR — Henry Gee's recent insinuation¹ that the authors of a very interesting reanalysis of the Vigilant *et al.*² data on mitochondrial Eve are "survivors of the Berkeley group (now all at PSU)" is not only in error but extremely distasteful. Those of us who were fortunate enough to work with Allan Wilson are more numerous than the few of us at Penn State (Maxson, Stoneking and Vigilant) and I can guarantee that none of us see ourselves as "survivors". On the con-

trary we are proud students and associates of a highly innovative scientist.

The Hedges *et al.* paper³ is not a rebuttal to Templeton⁴. Templeton showed that the original analysis was 'inadequate' and presented a more parsimonious tree that did not indicate an African root for the data. Hedges *et al.* examined more than 50,000 trees and then presented the single most likely tree generated by a neighbour-joining algorithm that was more appropriate for the dataset in question.

LINDA R. MAXSON

Institute of Molecular
Evolutionary Genetics,
Department of Biology,
Penn State University,
University Park,
Pennsylvania 16802, USA

1. Gee, H. *Nature* 355, 583 (1992).
2. Vigilant, L., Stoneking, M., Harpending, H., Hawkes, K. & Wilson, A. C. *Science* 253, 1503 (1991).
3. Hedges, S.B., Kumar, S., Tamura, K., & Stoneking, M. *Science* 255, 737 (1992).
4. Templeton, A. R. *Science* 255, 737 (1992).

Uninnovative Japan

SIR — I agree with Alexander Kennaway (*Nature* 355, 198; 1992) that being made to conform to bureaucratic budgets and targets handicaps good scientists, but he spoiled his argument by his references to Japan.

For a basic failing of the Japanese 'system' — one largely acknowledged by the Japanese themselves — is that it has not produced a fair share of the world's innovations.

Instead, the Japanese success may be due to the use of innovations by others and perhaps — as proposed by Michael E. Porter in *The Competitive Advantage of Nations* (Macmillan, London, 1990) — to fierce internal competition between manufacturers, initially behind a high tariff wall. Hardened by that struggle and the consequent need to pay close attention to the consumer, so the theory goes, Japanese manufacturers found the rest of the world easy.

Should this actually be the case — theories of industrial development are popular at the moment — then antimerger and antimonopoly laws would be far more important than increasing spending on research.

Research and development do have a place, of course, but it is believed that there is only a slight connection between the breakthroughs of pure science and local industry. Modern information flows are such that industry anywhere can pick up the breakthrough and pay rent for its use.

MARK LAWSON

Australian Financial Review,
273 Jones Street,
Ultimo, Sydney,
Australia 2007

Recovering Bacon's paradox

SIR — As a palaeontologist, I try to be especially sensitive to one great feature of human intellectual struggle: the incessant rediscovery of former and forgotten wisdom. *Plus ça change, plus c'est la même chose*. Zakaria Erzinclioglu is sorely distressed (*Nature* 355, 195; 1992) that we label fossils from the Earth's youth as ancient: "One of the most absurd and persistent misuses of words in science is the use of the word 'ancient' to describe species that flourished millions of years ago The animals themselves are not ancient in any sense at all. They died out a long time ago, when the Earth was young." Ralph Estling then replies (*Nature* 355, 667; 1992): "Trilobites and Aristotle are both; they were here early (earlier than we were), younger than we are, but from our habitual standpoint of looking back through time from where we happen to be at the moment, they are ancient and therefore older than we."

I do not dispute Estling's pluralistic resolution, but wish to point out that these current antagonists are replaying one of our oldest linguistic disputes, indeed a classic paradox of our literary traditions. Francis Bacon gave a first crisp statement in his *Advancement of Learning* in 1605 (though R. K. Merton, in his charming book, *On the Shoulders of Giants*, traces earlier uses to Giordano Bruno and even to the apocryphal book *II Esdras* of the *Vulgate*). Bacon wrote: *Antiquitas saeculi, juvenus mundi* (or, roughly, "the good old days were the world's youth") He later expanded in aphorism LXXXIV of the *Novum Organum*: "the old age of the world . . . is the attribute of our own times, not of that earlier age in which the ancients lived; and which, though in respect of us it was the elder, yet in respect of the world it was the younger." This observation was generally called "the Baconian paradox" when seventeenth century scholars so fiercely debated the relative merit of classical versus modern wisdom — a vital controversy surrounding the origin of modern science in Newton's age.

Jonathan Swift provided the most famous expression of Bacon's paradox in his wonderful satire on this great debate, *The Battle of the Books* (1704). One Friday night, in St James's Library, the books of Plato and Aristotle square off against those of Newton and Descartes:

Discord grew extremely high . . . Here a solitary ancient, squeezed up among a whole shelf of moderns, offered fairly to dispute the case, and to prove by manifest reason, that the priority was due to them But these [moderns] denied the premises, and seemed very much to wonder, how the ancients could pretend to

insist upon their antiquity, when it was so plain . . . that the moderns were much the more ancient of the two.

Bacon's paradox has since experienced cycles of oblivion and rediscovery. Jeremy Bentham, for example, left this aphorism among his unfinished papers (published posthumously in 1824): "What is the wisdom of the times called old? Is it the wisdom of gray hairs? No. — It is the wisdom of the cradle."

This antiquarian epitome would be worthless pedantry if the subject of Bacon's paradox did not raise two central questions for modern thought and the history of science. First, the issue does not arise until historicism and the notion of progress enter Western traditions — and these seminal ideas represent one of the great contributions of emerging science and technology. Why worry about whether the past was old or young if time runs in the repetitive cycles of Plato's Great Year and does not either advance or regress? The paradox has meaning only in the world constructed by science; indeed, Swift refers to Bacon's aphorism as "the modern paradox". Bacon emphasizes the same point in his obvious simile of time's passage with accumulation of knowledge in human ageing — his defence of the 'moderns' in the battle of the books: "And truly as we look for greater knowledge of human things and a riper judgement in the old man than in the young . . . so in like manner from our age . . . much more might fairly be expected than from the ancient times, inasmuch as it is a more advanced age of the world."

Second, Bacon's observation is a true paradox — a seemingly self-contradictory or absurd statement that happens to be true. As with other classical paradoxes, we both wince and revel in Bacon's idea because it embodies one of the inherent ambiguities of our complex lives, logics and phenomenologies. *Pace* Erzinclioglu, there is no solution; both perspectives, from both ends, are correct — and they do contradict one another. Rossini turned both 48 and 200 years old on 29 February 1992 (1800 and 1900 were not leap years in our Gregorian calendar). Frederick, the pirate apprentice, subject to the very same ambiguity (and faced with the prospect of bondage till the age of 88, his majority by true birthdays) touches the heart of our human predicament in W. S. Gilbert's sprightly words:

How quaint the ways of paradox!
At common sense she gaily mocks!

STEPHEN JAY GOULD
*Museum of Comparative Zoology,
Harvard University,
Cambridge, Massachusetts 02138, USA*

Babbage's dates

SIR — Heilbron and Bynum¹ mistakenly state that 1992 marks the bicentenary of the birth of Charles Babbage. In fact, Babbage was born on 26 December 1791, so that the chance for anniversary celebrations passed a week before publication of Heilbron and Bynum's article.

The confusion lies with Charles Babbage himself, who believed to the end of his life that he was precisely a year younger than he was. This was not known until rather more recently².

D. I. STEEL

*Anglo-Australian Observatory,
Coonabarabran, NSW 2357,
Australia*

1. Heilbron, J. L. & Bynum, W. F. *Nature* 355, 11–14 (1992).
2. Hyman, A. *Charles Babbage: Pioneer of the Computer* (Oxford University Press, 1982).

Insensitive ads

SIR — For several months, scientific journals have been publishing an advertisement for the Invitrogen company entitled "Native Expression" (for example *Nature* 23 January 1992, facing page 294). Although editorial staff are not responsible for the contents of advertisements, such an advertisement is insensitive, particularly at a time when the extermination of primitive people is a major concern. Taking DNA samples from them is good. To care about their horrible fate is much more important. Elementary human decency should forbid such a bad and vulgar demonstration of a poor and misplaced sense of humour.

This advertisement shows where the increasing desire to attract attention at any price can lead us.

ANDRÉ M. GOFFINET

*Facultés Universitaires
N-D de la Paix,
Département de Physiologie
Humaine,
B-5000 Namur,
Belgium*

Defining science

SIR — "Is molecular biology yet a science?" (*Nature* 355, 201; 1992) reminded me of a quotation from Lord Kelvin:

When you can measure what you are speaking about, and express it in numbers, you know something about it; but when you cannot measure it, when you cannot express it in numbers, your knowledge is of a meagre and unsatisfactory kind: it may be the beginning of knowledge, but you have scarcely, in your thoughts, advanced to the stage of science.

JANARDAN D. KHANDEKAR

*Division of Medical Oncology,
Evanston Hospital,
2650 Ridge Avenue,
Evanston, Illinois 60201, USA*

Polar path in geomagnetic reversals

S. K. Runcorn

A conducting layer at the core-mantle boundary under the Pacific Ocean could explain the tendency of the reversing geomagnetic pole to track along the Pacific rim.

THE earliest studied sequences of polarity reversals of the geomagnetic field in lava flows¹ and continental sediments² showed that the field reversed in a geologically very short time, estimated as a few thousand years. When normal and reversed remanent magnetizations were discovered in continuously deposited ocean bottom sediments³, it was hoped that the reversal process could be found, as this is essential in understanding the core dynamo⁴. Comparison of five records from different parts of the world seemed to show⁵ that the reversing field 0.7 Myr (million years) ago went through an axial quadrupole or octupole stage. But Laj and colleagues⁶⁻¹⁰ have recently found, in rapidly deposited sediments in Europe, several reversals over the past few million years in which the transition occurred by the dipole rotating through 180°. If the changing palaeomagnetic directions are plotted as north magnetic poles (virtual geomagnetic poles, or VGPs), assuming the transition fields were dipolar, they trace paths either across North and South America or 180° away in longitude near Australia and East Asia.

I suggest that a key to the explanation of this remarkable discovery is that these preferred polar paths bound the Pacific, where it has long been recognized that the geomagnetic secular variation is smaller than elsewhere¹¹. The reason for this may be that the varying core dynamo field is screened by a highly conducting layer in the lowermost mantle¹². Here I propose, as a simple model¹³, that this layer has metallic conductivity under the Pacific hemisphere and is insulating elsewhere. As the dipole begins to reverse along an arbitrary meridian, it induces currents in this hemispherical conducting shell and the torque rotates the core relative to the mantle until the reversing dipole path coincides with the two boundary-meridians in the Americas and eastern Asia.

Palaeomagnetic data

In two of the three middle-Miocene reversals studied in Zackinthus by Laj *et al.*⁶, the pole moved along the Asian path. In the Upper Miocene of Crete, Valet *et al.*⁷ have followed the transitions from reversed polarity (R) to normal (N) 7.25 Myr ago and, in three widely separated sites, 5.87 Myr ago. In these the VGP moved from south to

north across the Americas, whereas in the N to R transition 6.34 Myr ago found also in Crete, the VGP moved along the Asian path. The N to R transition at 5.67 Myr is anomalous, as the pole moved across the Atlantic. Detailed sampling of sediments of the Po Valley⁸ shows that during the Upper Olduvai event about 1.65 Myr ago, the VGP moved to southern latitudes, backtracked to the north and then reached the South Pole, all across the Americas. A marine core in the eastern Mediterranean⁹ also shows that during the Blake geomagnetic event (a short-lived 4,000-year reversal about 120,000 years ago, in the normal period which had existed since 0.7 Myr ago), the VGP moved southwards from the North Pole, backtracked, reached the South Pole and returned to mid-latitudes in the Southern Hemisphere, all across the Americas. It then moved rapidly across the Pacific (in about 10³ yr), and returned to the North Pole by the Asian preferred path.

I believe that these results alone show that the geomagnetic field reverses, not by disappearing or by becoming nondipolar, but normally through rotation of the dipole axis; if the field were nondipolar during reversal, the calculated poles would have no physical meaning and so their association with certain parts of the globe would be inexplicable. But the Blake event has also been followed from lacustrine sediments at Pringle Falls, Oregon¹⁴, showing a similar VGP path and, significantly, the same swing across the Pacific to end its return to N across the west Pacific. The Upper Olduvai path across the Americas has been confirmed from a mid-latitude deep-sea core taken in the Pacific Ocean¹⁵ and from continental sediments at Mar de Plata, Argentina¹⁶. The Upper Olduvai reversal has been observed¹⁷ in a core from the North Atlantic, but in this record part of the path lies across Australia and eastern Asia (not recorded in ref. 8). Finally, the paths of 25 reversals observed at 48 sites and reported in the literature are remarkably well confined to the two longitudes⁸.

Sceptics point to reversals that apparently do not fit into this pattern¹⁹⁻²¹. First, there is the reversal 0.7 Myr ago: the VGP paths from a mid-latitude site in the North Atlantic and from the southwest Indian Ocean lie through the Americas, whereas that from an equato-

rial Atlantic site lies 120° to the east through India²⁰. This is held to imply a nondipolar transition field, specifically an equatorial octupole²¹. Second, of the seven records of the lower Jaramillo (1.2 Myr) reversal⁸, five have paths across the Americas but two fall 90° further west. Such disagreements are regarded as evidence for nondipolar transitional fields, but this seems difficult to reconcile with the fact that many of the calculated paths do fall across the Americas, which, as I have argued above, is evidence that the transition field is dipolar. These anomalous records possibly need to be re-evaluated.

Secular variation

Early world magnetic surveys showed that both the geomagnetic secular variation and the nondipole field are smaller over the Pacific hemisphere than elsewhere¹¹. During this century several regions of secular change, in which the isoporic lines form clusters around the isoporic foci, where the rate of change of a magnetic field component is greatest, have been followed in their westward motion of 0.2° per year; but none lies in the Pacific¹⁸. As these isoporic foci were seen in the older maps to arise and decay on timescales of ~100 years, it is improbable that their absence this century in the Pacific is due to a statistical fluctuation. Our knowledge of the geomagnetic field in earlier centuries has recently been greatly expanded. We now know that the low secular variation over a region roughly coextensive with the Pacific has been a feature of the field since 1700 and, in particular, the westward drift is not evident²². Palaeomagnetic evidence, using the scatter of the directions of remanent magnetization in late Tertiary and Quaternary lava flows, also shows that the secular variation has been less over the Pacific than over the non-Pacific hemisphere²³⁻²⁵ during the past few tens of millions of years. The palaeomagnetic surveys, however, are limited in coverage and require further expansion²⁶.

In recent years, seismologists have found evidence for a solid heterogeneous layer at the base of the mantle (D'') in which the radial gradient of seismic velocities is small. Although this may be due to a large temperature gradient, it may also indicate substantial solid FeO, FeS and possibly iron, as well as silicate. I

make the hypothesis that beneath much of the Pacific hemisphere this layer has a nearly metallic conductivity, whereas elsewhere it has much smaller electrical conductivity. According to a well-known theorem of Maxwell, mixtures of metal and insulator can have a high or low electrical conductivity without change in relative amounts depending on their spatial arrangement. Screening by this conducting layer provides a reasonable quantitative explanation for the low secular variation observed in the Pacific. An infinite sheet of thickness d (cm) of electrical conductivity σ (c.g.s. e.m.u.) will reduce by $1/e$ a varying uniform magnetic field of period T , where $T = 4\pi^2 \sigma d^2$ (in seconds). The D'' layer is ~ 200 km thick, and if we take $\sigma = 10^3$ ohm $^{-1}$ cm $^{-1}$, a somewhat low metallic conductivity, $T = 500$ years. For the secular variation, harmonics higher than the first degree are prominent, and field changes of period ~ 100 years, often regarded as characteristic of the geomagnetic secular variation, will thus be markedly attenuated at the Earth's surface in the Pacific hemisphere but not elsewhere.

Support for this simple model comes from the map of long-wavelength S-wave velocity anomalies in the D'' layer where, over a region roughly below the Pacific, the velocities are $\sim 1\%$ low²⁷. This could be explained if this layer were 2% denser, with more solid FeO and FeS than elsewhere, as it is unlikely to be due to temperature effects. Anomalies of P-wave velocity of 1–3% in the D'' layer have also been observed, an extensive one under the Pacific²⁸. If this layer is 2% more dense, with more solid FeO and FeS than elsewhere, I suggest that it is also thinner here than elsewhere (in analogy with the oceanic versus continental crust). It follows from isostasy that the surface of the fluid core will be higher by 5 km. There seems to be seismological evidence for a greater core radius beneath the Pacific than elsewhere²⁹. At these pressures, FeO and FeS are likely to have electrical conductivities approaching that of a metal. Thus the simple model of hemispherical asymmetry in conductivity may be a good first approximation. There is also seismic evidence for considerable short-wavelength heterogeneity in the D'' layer³⁰, which could be caused by scatterers of scale 10–70 km with 1% velocity differences, entirely compatible with the 'Maxwell' model suggested above.

A recently discovered phenomenon of the core dynamo has an even shorter timescale. In the 'geomagnetic jerk', the second derivatives of the field changed within about a year³¹ at intervals of several decades. In the proposed model, changes on such a short timescale should be more completely screened by the

metallic shell. Magnetic observatories, on which this work is based, are sparse in the Pacific hemisphere, but the well-documented jerk of 1969 was not seen at Honolulu³² although it was seen in Guam near the edge of the shell. The maps of lines of equal discontinuity in each field component, or jerk contours, show that zero contours predominate over the Pacific hemisphere³².

An alternative explanation has, however, been suggested: that the dynamo generates less secular variation and nondipole field in the hemisphere of the core under the Pacific. The temperature differences associated with solid-state convection in the mantle, which has a timescale of 10–100 Myr, are supposed to diffuse into the core and exert a control over the core motions³³. Support for this suggestion has been adduced from the rough correlation between these two preferred paths and the broad longitudinal zones of high P-wave velocities, discovered by seismic tomography and interpreted as high-density and low-temperature anomalies³⁴. Suppose, although no detailed mechanism has been proposed, that these small zonal temperature anomalies diffuse into the core and somehow constrain the reversing dipole to follow one or other path. It would still remain unexplained why one hemisphere bounded by these anomalies, rather than the other, generates less secular variation and nondipole field. Moreover, I think that the most fundamental fact about the interaction of the core and mantle is that the torques coupling them are variable, for otherwise the irregular variations in the length of the day cannot be explained¹⁸: the core angular momentum changes so that the total angular momentum of the Earth remains constant, apart from the much slower deceleration due to tidal friction. The relative rotation of the core with respect to the mantle, measured by the westward drift of the nondipole parts of the geomagnetic field (at present 0.27° per year) shows a variation since 1820 roughly correlating with that of the length of the day. I conclude that the low secular variation under the Pacific cannot be attributed to some long-lasting hemispherical asymmetry of the core dynamo, for if there were such an asymmetry it would not remain under the Pacific for hundreds of years.

Significance has been attached¹⁰ to the fact that the two preferred paths of the transitional VGPs are also correlated with core velocity patterns calculated on certain assumptions. Theorists have attempted to derive information about the velocities in the core from geomagnetic secular variations³⁴. It has been assumed that the secular variation at the core–mantle boundary can be calculated from the magnetic field at the

Earth's surface: this, of course, is true only if electric currents between these surfaces are negligible. Magneto-hydrodynamic theorems are then used to derive certain aspects of the surface velocities in the core. The small core velocities found beneath the Pacific hemisphere, a striking feature of these calculations, lie between the preferred paths of the reversal VGPs³⁴, and this is held to support the view that one hemisphere of the core dynamo is producing less secular variation and nondipole field than the other because of (it is supposed) thermal coupling with the mantle. If, instead, the correct explanation of the low secular variation in the Pacific hemisphere is that electric currents flowing in the conducting hemispherical shell beneath the Pacific greatly modify it, then the assumption underlying the calculation of flow patterns in the core is not justified and this research will require not insignificant modification.

Rotation

What is the nature of the interaction between this hemispherical metallic shell, the D'' layer under the Pacific, and the reversing core field? If the core were alone in space, it is inconceivable that in successive reversals the VGP would follow the same path in a frame of reference rotating with the core: a fluid cannot have such a memory. Suppose therefore that the dipole initially begins to rotate in an arbitrary meridian such that the axial dipole component diminishes and an equatorial dipole grows. The component of the latter in the mid-meridian of the shell induces currents in it, which produce a nearly uniform field opposed to this component. The equatorial dipole experiences a torque which rotates it, and the core, about the Earth's axis until the dipole is rotating in the boundary-meridian of the shell, forcing the VGP path to lie either along the Americas or in eastern Asia. Changes of the dipole in this plane produce no torque about the Earth's axis. Palaeointensity determinations in sediments are uncertain, but the observations^{6–10} indicate that the dipole diminished during these reversals by factors of roughly 3–5. If the decreasing dipole strength outweighs the effect of rotation on the equatorial component, it decreases rather than increases: the induced field of the shell will reverse and tend to rotate the equatorial dipole into the mid-meridian. This fundamental property of electromagnetic induction (Lenz's law) could explain the rapid motion of the pole from one preferred path to the other, observed in the Blake records and perhaps in the Olduvai.

Quantitative discussion of this mechanism requires more precise information about timescales than is avail-

able at present. We know, however, that the Blake event took about 4,000 years. Thus the initial rotation of the dipole to mid-latitude in the Northern Hemisphere would have taken 100–1,000 years. The time taken for the pole to move across the Southern Hemisphere from the American to the Asian path must have been similar. This timescale is perhaps a factor of 2 or 3 longer than the predominant timescale in the secular variation. I estimated¹² that the hemispherical shell attenuates the secular variation by a factor of about 3, but the attenuation increases as the period of the varying field decreases and its harmonic degree increases. Therefore the reversing dipole, viewed from outside the shell, would have been less attenuated than the secular variation, by a factor of about 10–100. The uniform induced field inside the shell would therefore rise as the dipole rotates to a strength one or two orders of magnitude less than the dipole field at the core–mantle boundary, that is 0.5–0.05 G (S.K.R., manuscript in preparation). The torque on the dipole reaches a maximum when it is inclined at 45° to the Earth's axis and to the mid-meridian. If the present value of the Earth's dipole moment, 6.5×10^{25} G cm³, is used, the torque on the core can be as much as 10^{23} or 10^{24} dyne cm at maximum. Its mean value is likely to be an order of magnitude less.

It is tacitly assumed that the core rotates as a solid body during these processes, as in the interchanges of angular momentum between it and the mantle that give rise to the irregular changes in length of day. Generation of the geomagnetic field and its secular variations depends on toroidal velocity patterns, the radial variations of which are fundamental to the dynamo process, and poloidal (or convective) motions within the core. The secular variation and reversals reflect their changing relationships rather than any complete reorganization of core motions. Thus, when considering reversals and variations in length of day, it may be sufficient to treat the rotation of the core relative to the mantle as if it were a solid body carrying with it its magnetic field. The timescales are not so different that this assumption is unreasonable; 'solid-body rotation', on the fundamental principle of magnetohydrodynamics, would be expected as a first approximation.

The quantities involved may be appreciated by noting that the torque on the Earth produced by the lunar and solar tides, determined by satellite geodesy³⁵, is at present 4×10^{23} dyne cm. Allowance being made for a nontidal acceleration of the Earth, a striking confirmation and independent determination of this is that the total solar eclipse of 136 BC was observed in Baby-

lon, whereas had the Earth been rotating at a uniform rate the track of totality would have been through northern England³⁶. In the past 2,000 years, this decelerating torque has rotated the Earth through 57° relative to a uniformly rotating frame of reference. But the moment of inertia of the core is only about 12% of the whole Earth. It is reasonable that, although we do not know the time taken for the field to reverse or how the Earth's magnetic moment varies during reversal (it may diminish by a factor of 5), the core should rotate until the reversing dipole lies in the meridian plane 90° from the central meridian of the conducting shell. Rotating or changing in magnitude in this plane, the dipole experiences no torque.

Conclusion

I do not suppose that the D'' layer has been in its present state for more than a small fraction of the Earth's existence. It is thought that the inner body is pure iron solidifying from the core because of the cooling of the Earth; the lighter element, seriously thought to be oxygen, sulphur or silicon, is released and some rises to the core–mantle interface³⁷. The consequent release of gravitational energy is thought to be fundamental to field generation³⁸. Thus the D'' layer is a zone of chemical reactions⁴⁰ as well as the thermal boundary layer of mantle convection. Some geophysicists believe that plumes rise from it³⁹.

Experiments^{41,42} at high temperatures and pressure on minerals likely to be present in the lower mantle show that the electrical conductivity depends strongly on FeO content and that the

value chosen above is reasonable. Consequently, on the geological timescale, the layer will be in a state of flux chemically and thermally; the present, roughly hemispherical, asymmetry in electrical properties may have existed only for the past few 10^6 – 10^7 years. An association between the heterogeneities of the D'' layer and mantle convection might be expected, and the seismic tomography patterns in the lower, and even upper, mantle may be correlated with the distributions of other indicators of mantle motions, such as hotspots. It is therefore unlikely that the polar path during reversals would have been similarly constrained in earlier times as it apparently was in the late Cenozoic.

One excellent record exists of the transition pole path in the lower Triassic⁴³ which shows it following a great circle path, suggesting, in view of the data discussed above, that the field has normally reversed by the rotation of the dipole. This does not imply that the field always reversed in this manner even in the past 10 Myr; more palaeomagnetic data may indicate other modes of reversal. But the discovery¹⁰ of the path of the pole during reversals in recent geological times is consistent with what we know about the loose coupling between the core and the mantle, and seems to be explained by the simplest hypothesis accounting for the low geomagnetic secular variation in the Pacific hemisphere. □

S. K. Runcorn is at the University of Alaska, Fairbanks, Alaska 99775, USA and the Department of Physics, Imperial College of Science, Technology and Medicine, London SW7 2BZ, UK.

- Hospers, J. *Nature* **168**, 1111–1112 (1951).
- Irving, E. & Runcorn, S. K. *Phil. Trans. R. Soc.* **A250**, 83–99 (1957).
- Opdyke, N. D., Glass, B., Hays, J. D. & Foster, J. *Science* **154**, 349–357 (1966).
- Jacobs, J. A. *Reversals of the Earth's Magnetic Field* 230, (Hilger, Bristol, 1984).
- Hoffman, K. A. & Fuller, M. D. *Nature* **273**, 715–718 (1978).
- Laj, C., Guitton, S., Kessel, C. & Mazaud, A. *J. geophys. Res.* **93**, 11655–11666 (1988).
- Valet, J.-P., Laj, C. & Langeris, C. G. *J. geophys. Res.* **93**, 1131–1151 (1988).
- Tric, E. et al. *Phys. Earth planet. Int.* **65**, 319 (1991).
- Tric, E. et al. *Earth planet. Sci. Lett.* **102**, 1 (1991).
- Laj, C., Mazaud, A., Weeks, R., Fuller, M. & Herero-Bervera, E. *Nature* **351**, 447 (1991).
- Vestine, E. H., Laporte, L., Cooper, C., Lange, I. & Hendrix, W. P. *Carnegie Inst. Wash. Publ.* **578** (1947).
- Runcorn, S. K. in *Handbuch der Physik* **47**, 498–533 (Springer, Berlin, 1956).
- Runcorn, S. K. in *New Approaches in Geomagnetism and the Earth's Rotation* (ed. Flodmark, S.) **281** (World Scientific, Singapore, 1991).
- Herrero-Bervera, E. et al. *Phys. Earth Planet. Int.* **56**, 112–123 (1989).
- Herrero-Bervera, E. & Khan, M. A. *Geophys. J. Int.* **108**, 535–545 (1992).
- Ruocco, M. thesis, Univ. Stockholm (1990).
- Clement, D. M. & Kent, D. V. in *Rep Deep Sea Drilling Proj.* **94**, 831–852 (1986).
- Runcorn, S. K. *Trans. Am. geophys. Un.* **35**, 49 (1954).
- Bogue, S. *Nature* **351**, 445–446 (1991).
- Clement, B. & Kent, D. *Geophys. Res. Lett.* **18**, 81–84 (1991).
- Clement, B. M. *Earth planet. Sci. Lett.* **104**, 48 (1991).
- Bloxham, J., Gubbins, D. & Jackson, A. *Phil. Trans. R. Soc.* **A329**, 415–502 (1989).
- Doell, R. R. & Cox, A. *Science* **171**, 248–254 (1971).
- Bingham, D. K. & Stone, D. B. *Geophys. J. R. astr. Soc.* **28**, 337–355 (1972).
- McElhinny, M. W. & Merrill, R. T. *Rev. Geophys. Space Phys.* **13**, 687–708 (1975).
- Barton, C. E. in *Encyclopedia of Solid Earth Geophysics* (ed. James, D. E.) 560–577 (van Nostrand Reinhold, New York, 1989).
- Young, C. J. & Lay, T. *J. geophys. Res.* **95**, 17385–17402 (1990).
- Dziewonski, A. N. *J. geophys. Res.* **89**, 5929 (1984).
- Morelli, A. & Dziewonski, A. M. *Nature* **325**, 678 (1987).
- Bataille, K., Wu, R. S. & Flatté, S. M. *Pure appl. Geophys.* **132**, 1951 (1990).
- Courtillot, V., Ducruix, J. & le Mouél, J.-L. *C. R. Acad. Sci. Paris D287*, 1095–1098 (1978).
- le Mouél, J.-L., Ducruix, J. & Duijen, C. H. *Phys. Earth planet. Int.* **28**, 337–350 (1982).
- Gubbins, D. & Richards, M. *Geophys. Res. Lett.* **13**, 1521–1524 (1986).
- Bloxham, J. & Jackson, A. *Rev. Geophys.* **29**, 97 (1991).
- Lambeck, K. *The Geophysical Causes and Consequences of Earth's Variable Rotation* (Cambridge University Press, 1980).
- Muller, P. M. & Stephenson, F. R. in *Growth Rhythms and the History of the Earth's Rotation* (eds Rosenberg, G. D. & Runcorn, S. K.) 459 (Wiley, Chichester, 1975).
- Anderson, D. L. *Theory of the Earth* 366 (Blackwell, Oxford, 1980).
- Lowes, F. J. *Geophys. Surv.* **7**, 91–107 (1984).
- Loper, D. E. *Tectonophysics* **187**, 373–389 (1991).
- Knittle, E. & Jeanloz, R. *Science* **251**, 1438 (1991).
- Li, X. & Jeanloz, R. *J. geophys. Res.* **95**, 609 (1989).
- Li, X. & Jeanloz, R. *J. geophys. Res.* **96**, 6113 (1991).
- Herrero-Bervera, E. & Helsley, C. E. *J. geophys. Res.* **88**, 3406–3522 (1983).

Not so Great Attractor?

Recent cosmological surveys suggest a wholesale streaming motion of galaxies across the sky, conflicting with the idea that local galaxies move under the influence of a Great Attractor. But the issue is not settled.

FOUR years have passed since Donald Lynden-Bell and six colleagues described the concerted motion of about 400 elliptical galaxies towards the Centaurus region, in excess of their expected cosmological recession (*Astrophys. J.* **326**, 19; 1988). Later, Alan Dressler and Sandra Faber concluded that more distant galaxies in the same direction were receding less rapidly than cosmological expansion would require (*Astrophys. J. Lett.* **354**, 45; 1990). The conclusion seemed obvious, if surprising: a huge concentration of mass exists, sufficient locally to distort the Hubble expansion and to give galaxies peculiar motions of hundreds of kilometres a second. That was the Great Attractor.

Now, D. S. Mathewson, V. L. Ford and M. Buchhorn (*Astrophys. J. Lett.* **389**, 5; 1992) have published their own survey of galactic velocities and distances in the same patch of sky, but have come to a quite different conclusion: there is no Great Attractor. They say that the galaxies beyond the supposed Great Attractor are not falling back, but are moving away as rapidly as the foreground galaxies.

Adding to the confusion is a survey by J. Willick (*Astrophys. J. Lett.* **351**, 5; 1990) of galaxies in the Perseus-Pisces cluster, almost antipodean to the Great Attractor. Willick found that galaxies there are moving towards us, and so in the same direction as the wholesale motion found by Mathewson *et al.* In other words, all the galaxies from the Perseus-Pisces region on one side of us to the Centaurus region on the other are streaming across the Universe with a collective velocity of more than 500 km s^{-1} . If the Great Attractor does not exist, can there be an Even Greater Attractor further away still?

There is an even more bizarre interpretation. The cosmic microwave background has a substantial dipole moment, which is taken to mean that we are moving at about 600 km s^{-1} relative to a cosmic standard of rest in which the microwave background would be perfectly isotropic. It happens that this velocity is roughly consistent with the velocity at which Lynden-Bell *et al.* said we are moving towards the Great Attractor. This is tidy, meaning that the Great Attractor is more or less stationary in the cosmic rest frame, and that we and all the other galaxies in our neighbourhood are falling towards it.

But, if Mathewson *et al.* are right, that will no longer wash. Instead, they say that the apparent streaming motion of galaxies

(our own included) across the entire sky arises only because all the velocities have been referred to the supposed cosmic rest frame. So the data can be given quite a different meaning: if our galaxy is taken to be 'at rest', so are all the others, and the streaming disappears. But if we are not moving with respect to the microwave background, the microwave background must be moving with respect to us. It is possible to make cosmological models with an apparent anisotropic cosmic motion of this sort, but the average cosmologist would find them unappealing.

Fortunately (or unfortunately, for those who prefer the bizarre to the sensible), the original answer may be correct. Faber has pored over the Mathewson survey; she does not dispute the data, but is less than convinced by the interpretation. All such surveys must contend with the age-old problem of observational cosmology: redshifts (velocities) are easy to measure, distances are difficult. The redshift of any galaxy is composed of cosmological recession plus peculiar motion, and if the distance is independently known, the cosmological part of the redshift can be subtracted away, leaving only the peculiar motion.

There is, it happens, a fairly tight correlation between the absolute luminosity of galaxies and the velocity dispersion of stars within them, which can be measured by the width of a suitable emission line. Therefore, measure the velocity dispersion, estimate the absolute galactic luminosity from the known correlation, compare it with measured luminosity to get distance, figure out the cosmological redshift for that distance, subtract from the measured redshift, and find the peculiar velocity. By this procedure, peculiar velocities can be mapped out and attractors great and small, or large-scale streaming motions, show up.

If all galaxies were identical this would be easy. But the correlation between luminosity and velocity dispersion has some scatter, which translates into errors in distance determinations. This would still not be too bad, were it not for another pernicious effect: bias. Brighter galaxies are the more prominent, and so are the more likely to be included among the high redshift objects in deep galaxy surveys. Such a bias towards the higher intrinsic magnitudes is more than a nuisance, because distance determination itself depends on luminosity. So there is potential for all kinds of subtle systematic biases, and any bias that

can in principle occur will do so. So much has been known almost since observational cosmology began, but coping with bias still engenders argument.

Faber, speaking at a recent National Academy of Sciences colloquium at Irvine, California, said she believes that a proper accounting for bias would make the survey of Mathewson *et al.* consistent with her earlier work with Dressler, in which the far-side infall to the Great Attractor was established: Mathewson *et al.* provide a survey of galaxies in an 'ordinary' direction as well as towards the Great Attractor, and subtracting one from the other, to erase directly any systematic redshift-dependent effects, shows that the pattern of motion is indeed towards a relatively nearby centre, not an overall streaming.

In most galaxy surveys, independent distance estimates are unavailable, and only the redshifts are measured. In the absence of further information it would be impossible to separate the redshift into its cosmological and peculiar parts, but surveys also record the positions of galaxies. If peculiar velocities are due solely to the gravitational influence of galaxy clusters, then galaxy positions and velocities are related (by Poisson's equation for the newtonian gravitational potential), and with a sufficiently complete survey it becomes possible to analyse simultaneously the positions and redshifts of galaxies so as to reach a self-consistent picture yielding the peculiar velocities.

Michael Rowan-Robinson, discussing at Irvine such an analysis of the IRAS (Infra-Red Astronomy Satellite) sky survey, professed himself satisfied that all is more or less well. The most prominent large-scale peculiar galactic motions are reasonably well accounted for by the largest galaxy clusters. But one discrepancy stands out: the region between us and Perseus-Pisces is relatively empty, which would suggest that galaxies in the Perseus-Pisces region ought to be moving away from us. In fact, as Willick found, they are moving towards us. Rowan-Robinson thinks this will be resolved with more comprehensive surveys, but, in the light of Mathewson *et al.*, there is a lingering suspicion that something funny may be happening along the Perseus-Pisces/Centaurus axis. A few more years should tell us whether the Universe is really uniform and isotropic, in the large, or if relativists ought to start dusting off their anisotropic cosmologies.

David Lindley

Threats from debris avalanches

Lee Siebert

THE 1980 eruption of Mount St Helens was one of the relatively few volcanic eruptions that radically altered scientific thought about how volcanoes behave. The massive landslide that accompanied it and removed more than one kilometre of the volcano's summit focused attention on the instability of volcanic edifices. And the spectacular lateral blast that devastated 600 km² of Pacific Northwest forest highlighted a previously little-known eruptive process. The ensuing dozen years have seen a growing recognition that volcanoes are prone to sudden collapse, with consequences that extend far beyond the volcano. Begét and Kienle's study of volcanic landslide deposits at Mount St Augustine in Alaska, on page 701 of this issue¹, reveals a previously undocumented pattern of episodic collapse and regrowth at a single volcano.

Although the hazards of erupting volcanoes are readily apparent, volcanoes themselves have an air of permanence that belies their inherent instability. Under certain conditions, a volcanic edifice can fail catastrophically, producing extremely mobile, gravity-driven debris avalanches that accelerate rapidly down-slope and can travel at velocities estimated to be as large as 100 m s⁻¹ (360 km hr⁻¹) for distances that can exceed ten times the vertical drop. Large-volume avalanches have travelled many tens of kilometres over regional slopes of less than 1°, depositing volcanic debris tens of metres thick over areas of hundreds to more than a thousand square kilometres.

The potential scale and mobility of volcanic landslides is highlighted by a newly published study of a late-Pleistocene debris-avalanche deposit in Mexico². Collapse of Nevado Colima volcano produced an avalanche that covered an estimated 2,200 km² area and travelled 120 km to the Pacific Ocean, the Earth's longest currently-known run-out of a debris avalanche (submarine runouts excepted). Catastrophic structural failure of a volcanic edifice is not restricted to steep-sided subduction-zone volcanoes. The ocean floor bordering the Hawaiian islands contains even more massive Neogene to Quaternary debris-avalanche deposits that can exceed a thousand cubic kilometres in volume resulting from collapse of low-angle shield volcanoes³.

These unusual deposits were long considered to be produced by mudflows originating from catastrophic draining of crater lakes. In the few years preceding the Mount St Helens eruption, volcan-

ologists in Japan began noting that these deposits differ texturally from mudflow deposits and resemble those of non-volcanic landslides. The Mount St Helens eruption marked the first time that edifice collapse and progress of a large volcanic debris avalanche were documented scientifically as they occurred.

The most comprehensive examination of a large debris-avalanche deposit to date, a study at Mount St Helens by Harry Glicken⁴, who lost his life at Unzen volcano in Japan last year,



A landslide from the collapse of the summit of Alaska's Mount St Augustine volcano formed foreground hills. Nearly a dozen such events have occurred here in the past 2,000 years.

helped define the characteristics of these deposits. Re-examination of similar deposits at volcanoes worldwide has revealed that a process once thought to be extremely rare is not all that uncommon. Debris-avalanche deposits have now been identified at several hundred volcanoes; edifice failure is seen to have occurred globally an average of four times a century during the past 500 years. The potential for catastrophic failure of volcanoes and attendant hazards implications are now recognized, but the precise mechanisms for the extreme mobility of volcanic debris avalanches remain a topic of debate.

Mount St Augustine volcano in Alaska's Cook Inlet has been the site of numerous eruptions like that at Mount St Helens, most recently in 1883, when collapse of the summit produced a debris avalanche which itself created large tsunami waves as it entered the sea⁵. Begét and Kienle's island-wide stratigraphic study¹, constrained by radiocarbon dates from interlayered volcanic deposits and soils, has uncovered evidence for at least eleven large debris avalanches at Augustine Island during the past 2,000 years, suggesting an unprecedented collapse recurrence interval of about 150–200 years. Begét and Kienle calculate that the edifice is rebuilt by an influx of lava at a rate of 1–3 million cubic metres a

year, among the highest rates found at subduction-zone volcanoes⁶. Only at Egmont volcano in New Zealand is a comparable number of collapse events known, but the average recurrence interval at that larger volcano is measured in many millennia rather than centuries⁷.

Cyclic collapse and regrowth at Mount St Augustine has produced an unusual volcano whose profile resembles that of a classic strato-volcano (see figure), but that actually consists of a summit lava dome complex surrounded by a ring of distant debris-avalanche deposits resulting from collapse of the volcano and closer block-and-ash flow deposits associated with its regrowth.

In addition to the direct hazards — debris avalanches, lateral blasts and

mudflows — collapse of coastal volcanoes can produce tsunamis when avalanches sweep into the sea, affecting areas far from the volcano. Ten-metre-high tsunami waves damaged villages around Cook Inlet when Mount St Augustine collapsed in 1883. A century later, the volcano has rebuilt itself to its pre-failure height. Although the next eruption will not necessarily involve failure of the edifice, the collapse recurrence rates determined by Begét and Kienle indicate that this possibility should be considered during any future eruption. The 1883 tsunami fortuitously occurred at low tide at a time when the southern Cook Inlet was sparsely populated; a similar event today could have more severe consequences. Modelling of avalanche mobility suggests that any characteristic avalanche at Mount St Augustine would reach the coast, producing a tsunami⁵. Tsunami travel-time modelling suggests that only about an hour's warning would be available for towns in Cook Inlet following edifice collapse⁸.

The Alaska Volcano Observatory — jointly operated by the US Geological Survey, the University of Alaska and Alaska's state geological survey — has installed deformation and seismic monitoring networks on Augustine Island that will help anticipate future eruptions. Great strides have been made in volcanic

monitoring techniques during the dozen years since Mount St Helens erupted, and eruption prediction is fast becoming a reality at well-instrumented volcanoes. Political memories are short, however, and major budget cuts have been proposed for the US Geological Survey's Volcano Hazards Program at a time when potentially hazardous volcanoes in populated regions of Alaska and the western United States are insufficiently instrumented and lack even basic geological mapping.

Increased global population has placed large numbers of people within range of major volcanic landslides and associated eruptive phenomena. Although it is sometimes possible to anticipate the run-out and azimuth of large volcanic landslides, the causes of destabilization of volcanoes are complex, and precursors highly variable. The collapse of Mount St Helens was preceded by phreatic (steam-driven) eruptions and extensive deformation related to high-level magmatic intrusion. Subsequent calcu-

lations⁹ showed that this would have permitted a rough prediction of the time of failure. But collapse at other volcanoes has occurred without magmatic activity and even in the absence of eruptions. Disciplines as distinct as materials science, fluid mechanics, seismology and geodesy must be brought together if such catastrophic failures are to be predicted. □

Lee Siebert is at the Global Volcanism Program, Smithsonian Institution, Washington, DC 20560, USA.

1. Begét, J. E. & Kienle, J. *Nature* **356**, 701–704 (1992).
2. Stoopes, G. R. & Sheridan, M. F. *Geology* **20**, 299–302 (1992).
3. Moore, J. G. et al. *J. geophys. Res.* **94**, 17,465–17,484 (1989).
4. Glicken, H. thesis, Univ. California at Santa Barbara (1986).
5. Siebert, L., Glicken, H. & Kienle, J. *Nat. geogr. Res.* **5**, 232–249 (1989).
6. Crisp, J. J. *Volc. Geoth. Res.* **20**, 177–211 (1984).
7. Palmer, B., Alloway, B. F. & Neall, V. E. in *Sedimentation in Volcanic Settings* (eds Fisher, R. V. & Smith, G. A.) *Soc. Econ. Paleont. Min. Spec. Publ.* **45**, 89–98 (1991).
8. Kienle, J., Kowalik, Z. & Murty, T. S. *Science* **236**, 1442–1447 (1987).
9. Voight, B. *Nature* **332**, 125–130 (1988).

EVOLUTIONARY GENETICS

Selfish genes move sideways

Gregory D. D. Hurst, Laurence D. Hurst and Michael E. N. Majerus

A WIDE variety of ultra-selfish genetic elements dwell within the genomes of certain organisms — elements which, for instance, disrupt meiosis and distort sex allocation, and generally propagate themselves at a cost to the host genome¹. Their distribution has traditionally been considered to be determined largely by phylogeny; that is, if your ancestor had a selfish element, you may too. Work by O'Neill and colleagues on the evolutionary biology of the selfish elements causing cytoplasmic incompatibility², and by Maruyama and Hartl³ on the evolutionary biology of the mariner transposable element, means this view must now be changed. Far from being stuck within a phylogenetic lineage, ultra-selfish elements seem to have moved horizontally into previously uninfected lineages.

The selfish behaviour of cytoplasmic genes, as revealed by cytoplasmic incompatibility, was first described in the mosquito *Culex pipiens* and has since been recorded in several species⁴. Within any population of *C. pipiens*, there are two sorts of individuals: those that bear bacteria of the genus *Wolbachia* and those that do not. These cytoplasmic symbionts are found in eggs and subsequent zygotes, but not in mature sperm. When *Wolbachia*-infected males mate with uninfected females, the eggs die. All other crosses are viable. This phenomenon has been interpreted as the result of the selfish behaviour of the vertically inher-

ited bacteria: a case of intragenomic conflict⁴.

O'Neill *et al.*² selectively amplified, cloned and sequenced the 16S ribosomal RNA genes of cytoplasmic elements from the ovaries of insects of six different species, as well as those of two different lines of *Drosophila simulans*. The sequences are highly conserved, and the authors therefore argue that all of the symbionts should be classified as members of the same 'species' and that horizontal transmission has occurred. Cytoplasmic incompatibility is thus not a general physiological reaction to infection by a diverse range of bacteria, but a specialized response by the same bacteria in a wide variety of species². Whether the agent responsible for cytoplasmic incompatibility in other species (isopod crustaceans⁵ for example) is the same as that found by O'Neill *et al.*, has yet to be investigated.

O'Neill *et al.* went on to look at the genomes of species where incompatibility is not known, and found *Wolbachia*. The question of whether *Wolbachia* in these species exist through an advantage accruing from incompatibility, or are maintained in other ways, remains open. But the presence of this element in many insect species means that it will probably be found elsewhere.

The notion that horizontal transmission of the element has occurred is supported by comparisons of the phylogeny of the *Wolbachia* elements found

in different species, determined by 16S rRNA sequence divergence, and the phylogeny of the host². The two do not correspond. A similar study concentrating on the phylogeny of *Wolbachia* (also determined by 16S rRNA sequences) within eight different lines of *D. simulans* (phylogeny determined through mitochondrial sequence divergence) reaches the same conclusion: host phylogeny does not reflect *Wolbachia* phylogeny⁶. This inconsistency of host and element phylogeny corroborates the thesis that horizontal transmission has occurred. But errors may easily be made in constructing phylogenies from single DNA sequences⁷, and element phylogenies will have to be confirmed by examination of an independent sequence, such as that of the 23S rRNA gene.

Transmission between species may be a common property of ultra-selfish genes, of which cytoplasmic incompatibility agents are but one class. In particular, transposable elements (genetic elements which spread by being able to duplicate and reinsert themselves into the genome) seem to show some degree of horizontal transmission. The best studied transposable elements are those in drosophilids; a variety of them have been characterized within the fruit flies, and many of them occur in more than one species.

From Maruyama and Hartl's study³ of the mariner element, it seems that horizontal transmission is partly responsible for the element's distribution. Mariner occurs widely in the *D. melanogaster* species group, and was also found in the distantly related genus, *Zaprionus*. Sequencing of the *Zaprionus* mariner element revealed greater similarity between it and the element found in some members of the *melanogaster* group than between mariner elements within the *melanogaster* group. The sequence of alcohol dehydrogenase genes in the drosophilids confirmed that *Zaprionus* is a distant relative of the *melanogaster* group: phylogeny alone cannot adequately explain the occurrence of mariner in the *Zaprionus* genus, but phylogeny combined with occasional inter-specific horizontal transfer can³. A study comparing the phylogeny of another transposable element, the P-element, with that of its hosts reached a similar conclusion⁸.

If horizontal transfer of ultra-selfish genes does occur, its incidence will depend on three factors. First, on an ecological link with a species bearing the agent. Transfer could occur when an infected parasitoid species uses an uninfected host species⁷, but more generally it might happen when different parasites share a common host species or different hosts are used by the same parasite

species; in this latter category, it has been suggested that a shared mite species is a vector for the P-element⁹, and that shared viruses are carriers of other transposable elements¹⁰. Other links operating through predation and absorption through the gut may also be important.

The second criterion for establishment of selfish elements in a new host species is that the strategy used by the element must work in the new host. The cytoplasmic incompatibility organism will not necessarily be able to destroy uninfected eggs on fertilization, and may not be able to undergo vertical transmission in infected individuals. Transformation of uninfected species with transposable elements from closely related groups may be limited by the ability of the element to transpose¹¹, although artificial intra-generic transformations do work¹². Lastly, the element must increase in the population by virtue of its ultra-selfish behaviour. This consideration is, to date, the most studied.

Is a history of horizontal transfer a common property of ultra-selfish genes? To answer that question studies will have to be carried out on other groups of cytoplasmic selfish genes, such as those causing a bias in resource allocation towards female progeny. One interesting question is whether horizontal transfer is necessary for the survival of these elements over evolutionary time. It certainly is one explanation for the continual existence of these elements in spite of continual selection for resistance by their hosts. If ultra-selfish genes that cannot move between organisms are more likely to become extinct, a predisposition to horizontal transmission would be evident in extant elements. □

Gregory D. D. Hurst and Michael E. N. Majerus are in the Department of Genetics, University of Cambridge, Downing Street, Cambridge CB2 3EH, UK. Laurence D. Hurst is in the Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, UK.

1. Werren, J. H., Nur, U. & Wu, C.-I. *Trends Ecol. Evol.* **3**, 297–302 (1988).
2. O'Neill, S. L., Giordano, R., Colbert, A. M. E., Karr, T. L. & Robertson, H. M. *Proc. natn. Acad. Sci. U.S.A.* **89**, 2699–2702 (1992).
3. Maruyama, K. & Hartl, D. J. *molec. Evol.* **33**, 514–524 (1991).
4. Hurst, L. D. *J. theor. Biol.* **148**, 269–277 (1991).
5. Legrand, J.-J. & Juchaut, P. *Boll. Zool.* **53**, 161–172 (1986).
6. Rousset, F., Vautrin, D. & Solignac, M. *Proc. R. Soc. Lond. B* **247**, 163–168 (1992).
7. Breeuwer, J. A. J. *et al. Insect molec. Biol.* (in the press).
8. Daniels, S. B., Peterson, K. R., Strausburgh, L. D., Kidwell, M. G. & Chovnick, A. *Genetics* **124**, 339–355 (1990).
9. Houck, M. A., Clark, J. B., Peterson, K. R. & Kidwell, M. G. *Science* **253**, 1125–1129 (1991).
10. Miller, D. W. & Miller, L. K. *Nature* **299**, 562–564 (1982).
11. O'Brochta, D. A. & Handler, A. M. *Proc. natn. Acad. Sci. U.S.A.* **85**, 6052–6056 (1988).
12. Daniels, S. B., Chovnick, A. & Kidwell, M. G. *Genetics* **121**, 281–291 (1989).

Dating supernova remnants

R. N. Manchester

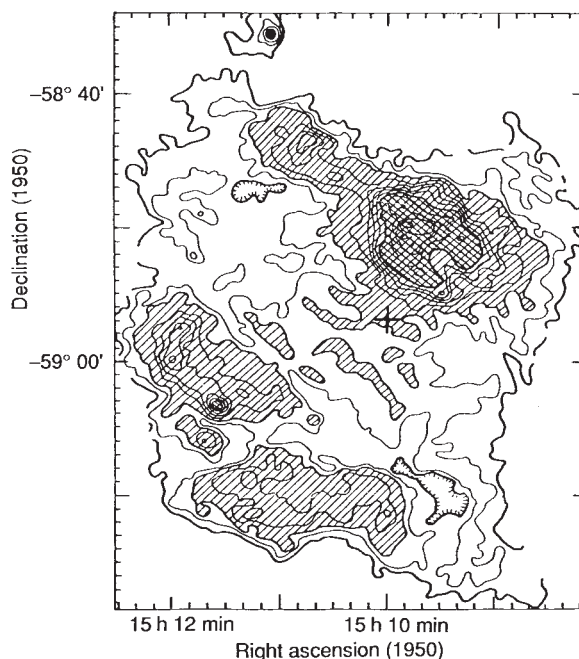
ESTIMATING the ages of supernova remnants is a difficult business. These expanding shells of emission are formed by the interaction with the surrounding interstellar gas of high-velocity debris from the supernova explosion and are normally most prominent at radio wavelengths. Evolutionary models can be used to estimate the ages, but these estimates are not very reliable. Sometimes a pulsar, the collapsed core of the exploded star, can be associated with the supernova remnant and the age estimated from the pulsar parameters. By far the best way of determining the age, though, is to observe the supernova itself. But supernovae are rather rare events — SN1987A in the Large Magellanic Cloud was the first supernova visible to the unaided eye for nearly 400 years.

Fortunately, transient celestial phenomena were taken very seriously by ancient Oriental civilizations — often they were considered a good and sufficient reason to lop off the heads of the opposition — and accurate records of these events were kept. One of the earliest such records to survive is of the supernova of AD 185. This supernova occurred in the southern Milky Way where there is a relatively high concentration of supernova remnants so that its identification is somewhat uncertain. On page 690 of this issue¹, S. E. Thorsett argues that this 'guest star' created the pulsar PSR1509–58 and the associated remnant MSH15–52.

Unlike the supernova of AD 1054, now identified with the Crab Nebula, which was documented by five independent far-eastern observers, only one record of the AD 185 event survives. This is no doubt partly because of the ravages of time, but also because of the far southern declination of this supernova. It rose only a few degrees above the horizon at the latitudes of the Chinese observatories. Fortunately, the existing record is sufficiently detailed to give an estimate of the supernova position on the sky and hence to permit identification with a present-day supernova remnant. Clark and Stephenson, in a pioneering study², suggested that the supernova should be identified with the remnant MSH14–63.

There is a problem with this identification though, in that a supernova at this position would have been lost in twilight in the last month when it was reported to be visible.

The radio remnant MSH15–52, also known as G320.4–1.2 from its galactic coordinates, is somewhat further from the best estimate of the supernova position than is MSH14–63. However, there is a pulsar, PSR1509–58, located near the centre of this remnant (see figure) and the observed rate of slowdown of the pulsation frequency shows³ that this pulsar is young, only about 1,700 years.



A long-stay guest? Map of the southern supernova remnant MSH15–52 at a radio frequency of 1,415 MHz (from ref. 9). Position of the pulsar PSR1509–58 marked by a cross.

The near coincidence of this age with that of the supernova of AD 185 strongly suggests that PSR1509–58 was born in this supernova explosion. Such events are expected only every 50–100 years in our Galaxy and to have two occurring in nearly the same direction within a few decades is very unlikely. Thorsett points out that a supernova at the position of PSR1509–58 would have been visible at night during the period of reported observation, thus removing the discrepancy with the historical record.

Because of the spatial coincidence of PSR1509–58 with MSH15–52 and the relative youth of the pulsar, the two were immediately seen as associated when the pulsar was discovered⁴. However, this association is not without controversy. The age of a supernova remnant can be estimated from its ex-

pansion velocity. Measurements of this type⁵ suggest that the remnant is about 10,000 years old, an order of magnitude greater than the estimated age of the pulsar. This result is consistent with less-direct arguments based on the size and surface brightness of the radio remnant⁶.

Evolutionary models for supernova remnants are used to interpret both the velocity and surface brightness results. However, problems arise because these models assume that the remnant expands into a uniform medium, whereas the interstellar medium in the vicinity of supernova remnants is far from uniform. Indeed, it is quite possible that fast winds from the progenitor star before it exploded (or from nearby similar stars) could have largely evacuated a bubble in the surrounding interstellar medium. In this case, the ejecta from the exploded star would have expanded freely before encountering the boundary of the bubble and generating the observed radio and optical emission. If this happened, the true age of the remnant would be much less than that derived from the models.

Pulsar ages are also somewhat model dependent, but it seems unlikely that their true age is much greater than the age indicated by the rate at which their spin is slowing down. That would require that the pulsar magnetic field increases with time. Just such an evolution has been suggested on theoretical grounds⁷, but the strong fields observed in the Crab pulsar and another young pulsar associated with a supernova remnant in the Large Magellanic Cloud⁸, PSR0540-69, do not support the idea.

Identification of PSR1509-58 and MSH15-52 as remnants of the supernova of AD185 would fix their age and would clearly be of great value in studies of pulsar and supernova-remnant evolution. It would be only the second pulsar (after the Crab) to be associated with an observed supernova. Thorsett makes a good, but perhaps not compelling case for the identification. I suspect the controversy is not yet over. □

R. N. Manchester is at the Australia Telescope National Facility, CSIRO, PO Box 76, Epping, New South Wales 2121, Australia.

Age and the unisexual lineage

John Maynard Smith

WITHIN the space of a few months, three groups have published work exploiting the peculiar biology of unisexual fish and amphibia. Two of the papers (by Spolsky *et al.*¹ and Hedges *et al.*²) appear on pages 706 and 708 of this issue; the other (by Quattro *et al.*³) was published in *Proceedings of the National Academy of Sciences*. The object of the studies is to draw conclusions about the histories of the nuclear and mitochondrial components of the genome, and about the evolutionary longevity of clones.

There seem to be no fully unisexual (parthenogenetic) fish or amphibia. All-female populations do exist, but their eggs only develop after stimulation by a sperm from a male of a sympatric sexual species. In some cases, the male contributes no chromosomes to the new individual, which is genetically identical to her mother. In other cases of so-called hemiclinal inheritance, the male provides a complete haploid genome, which is expressed in the offspring but is eliminated before meiosis, without recombination with the maternal genome. In this respect, anamniote vertebrates differ from lizards, of which there are fully parthenogenetic populations, able to reproduce without males.

Hedges *et al.*² studied salamanders of the genus *Ambystoma*. There are four sympatric bisexual species in eastern North America, and many unisexual populations, ranging from diploid to pentaploid. Electrophoretic studies of nuclear genes show that the unisexual populations are hybrids, with genes from two, or sometimes three, of the sexual species. Their eggs must be stimulated by a sperm from one of the sexual species. Typically, the chromosomes in the sperm are eliminated. However, the association of a wide range of nuclear genotypes with rather uniform mitochondria (see below) shows that a male genome is occasionally incorporated, giving rise to a novel clone, albeit one with the same mitochondrial genome.

The authors sequenced a 307-base-pair region of the mitochondrial cytochrome *b* gene from 46 salamanders — 20 of the unisexual hybrids and 25 of the four sexual species, with a different genus, *Plethodon*, as an outgroup. Eighteen of the 20 hybrids form a separate monophyletic lineage within the phylogenetic tree. Hedges *et al.* conclude that the mitochondrial (mt) DNA of the hybrids has been derived from a distant ancestor, now absent from the region and perhaps extinct, and has been transmitted in a clonal fashion, whereas the hybrids are continually acquiring nuclear

DNA from the sexual species. Assuming a cytochrome *b* sequence divergence rate of 2.5 per cent per million years as observed in other vertebrates, they estimate that the origin of the unisexual lineages occurred about 4 million years ago. The most striking conclusion to be drawn is that the mitochondrial and nuclear components of the genome have quite different histories.

Similar conclusions were reached by Spolsky *et al.*¹, who carried out a restriction-site analysis of mtDNA from 48 clones and 18 sexual individuals of *Ambystoma*. They found only two similar haplotypes among the clones, but much greater variability within sexual species. They also conclude that mtDNA in the clones is ancient — 5 million years old, they estimate.

In their paper, Quattro *et al.*³ describe studies on unisexual populations of a fish, *Poeciliopsis*, in a series of Mexican streams. These fish are hemiclinal. The populations analysed derive their permanent maternal genome from the sexual species *P. monacha*, and their temporary paternal genome from *P. occidentalis*. The two sexual species coexist in the Rio Mayo. In this stream, new unisexual hybrid lineages are repeatedly formed: they are ecologically diverse. In four streams to the north, only one of the sexual species, *P. occidentalis*, is found. There is also a rather rare unisexual population. Tissue grafting, and electrophoretic analysis of nuclear genes, show that these northern unisexual fish belong to a single hemiclone, which presumably arose from a single hybridization event in the Rio Mayo and has subsequently colonized the northern rivers in a stepwise manner.

Quattro *et al.* analysed restriction-site variation in mtDNA in the two sexual species and in the unisexual hybrids. As expected, the mtDNA of the northern hemiclone resembles that of *P. monacha*. The average sequence divergence between members of this hemiclone is 0.12 per cent, with a maximum value of 0.30 per cent. This is less than the divergence within the two sexual species, and much less than the divergence (11 per cent) between them. The authors estimate that this northern hemiclone is at least 100,000 years old.

The problem of the age of clones is important for theories about the evolution of sex and recombination. What Quattro *et al.* call the "conventional view"⁴ is that clones are short-lived on an evolutionary timescale. I do not think that this view is contradicted by the new results. Hedges *et al.* say that the hybrids

1. Thorsett, S. E. *Nature* **356**, 690-691 (1992).
2. Clark, D. H. & Stephenson, F. R. *The Historical Supernovae* (Pergamon, Oxford 1977).
3. Manchester, R. N., Durdin, J. M. & Newton, L. M. *Nature* **313**, 374-376 (1985).
4. Seward, F. & Harnden, F. R. *Astrophys. J.* **256**, L45-L47 (1982).
5. van den Bergh, S. & Kamper, K. W. *Astrophys. J.* **280**, L51-L54 (1984).
6. Seward, F., Harnden, F. R., Murdin, P. & Clark, D. H. *Astrophys. J.* **267**, 698-710 (1983).
7. Blandford, R. D., Applegate, J. H. & Hernquist, L. *Mon. Not. R. astr. Soc.* **204**, 1025-1048 (1983).
8. Seward, F., Harnden, F. R. & Helfand, D. J. *Astrophys. J.* **287**, L19-L22 (1984).
9. Caswell, J. L., Milne, D. K. & Wellington, K. J. *Mon. Not. R. astr. Soc.* **195**, 89-99 (1981).

they looked at "... represent the most ancient unisexual lineage known". But they also argue that the nuclear genes in the lineage are continually being replaced by genes from the four sexual species. The only component of the genome that has existed, with little or no recombination, for 4 million years, is the mtDNA. But if, as seems likely (and as is consistent with the data from these three papers), mtDNA is uniparentally inherited, it has existed with little recombination for 10^9 years.

Hedges *et al.* do not argue that their data have any special significance for the evolution of recombination: their emphasis is on the independent histories of nuclear and mitochondrial DNA. Spolsky *et al.* say that "such long-term persistence raises a question about the necessity of sexual reproduction ... to lineage survival", but they conclude (reasonably in my view) that rare replacement of nuclear genomes may have contributed to the long-term survival of clonal salamander lineages.

Quattro *et al.*, however, were specifically trying to estimate the age of a clone, because of its relevance to theories of sex. They regard their data as evidence against the view that clones are short-lived, and conclude that "the rarity of unisexual taxa might have as much to do with low origination rates as with high extinction probabilities". But, in evolutionary terms, 100,000 years is but an evening gone: it is important to know that a female genome of *Poeciliopsis* has existed without recombination for that long, but it does not contradict the conventional wisdom. The idea that low origination rates explain the rarity of parthenogenesis can at best be only a partial explanation. After all, sex probably originated just once, whereas parthenogenetic strains have arisen from sexual ones many times. If origination rates determined the relative frequencies of the two types of reproduction, we would all be parthenogens.

There are, however, some unisexual populations which seem to be genuinely old in evolutionary terms. The most dramatic are the bdelloid rotifers, in which no one has ever seen a male. Have these rotifers invented an alternative to males as a means of exchanging genetic material? If not, how do they manage without recombination? □

John Maynard Smith is in the School of Biological Sciences, University of Sussex, Brighton BN1 9QG, UK.

A mantle thermometer

Peter Shearer

DENSE regional networks of seismographs in the western United States, originally deployed to study local earthquakes, are now being used to reveal details in the structure of subduction zones thousands of kilometres away. Using array processing techniques common in exploration geophysics, Vidale and Benz on page 678 of this issue¹ combine records of distant deep earthquakes recorded on hundreds of individual seismometers to image faint near-source reflections from seismic discontinuities in the upper mantle near subducting slabs. The depth of these discontinuities can be used as a thermometer for rocks hundreds of kilometres below the surface, to show how they are affected by the arrival of cold oceanic lithosphere sinking into the hot mantle.

For many years, the US Geological Survey and other organizations have operated regional networks of hundreds of seismographs in California, Nevada, Washington and Utah. Data from these networks are used to locate and study local earthquakes in these active seismic areas. However, data from more distant events around the world (teleseisms) have been pretty much overlooked. Vidale and Benz have collected and combined data from five different seismic networks to form a single array of 881 stations spanning over 1,000 km. This was a formidable organizational task, as the networks store their data separately in a variety of formats.

Vidale and Benz set out to use the arrays to obtain a high-resolution picture of the internal structure of the Earth's slowly circulating mantle. The great advantage of using such a large number of stations is that data 'stacking' techniques can be used to suppress the effects of incoherent noise and increase the visibility of weak seismic phases such as those resulting from reflections off upper-mantle discontinuities. The seismic waves, both fast compressional (P) and slower shear (S) waves, were from a number of deep earthquakes from subduction zones around the Pacific basin, where cold oceanic lithosphere sinks into the mantle beneath the surrounding plates. For each earthquake, Vidale and Benz consider waves that arrive after the first, direct P-wave arrivals and before the later, surface-reflected pP-wave arrivals (Fig. 1). By combining the results from hundreds of individual recordings, they identify waves converted from

the S to the P phase on transmission through a discontinuity at a depth of 685 km near the source. They also recognize P-wave reflections from the underside of discontinuities at depths of 210 and 400 km (Fig. 2). The closeness of two of these to the depths of known global discontinuities^{2,3} at 660 km and 410 km (more accurately, 415 km) suggests that the presence of the subducting slab depresses the former by about 25 km and elevates the latter by about 15 km.

S-to-P converted arrivals have been identified in previous studies of subduction zone events⁴⁻⁶, but the underside P reflections had not previously been observed. Our global maps⁷ of the 660-km discontinuity obtained from long-period shear-waves indicate regional depressions of about 20-km deep near subduction zones, but these maps have limited horizontal resolution (about 1,000 km). Vidale and Benz's data have

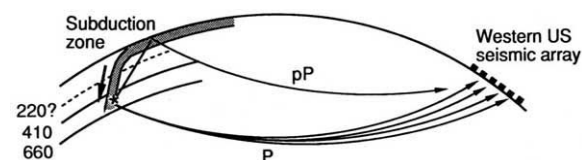


FIG. 1 Vidale and Benz study seismic waves from deep earthquakes generated as cold, dense oceanic lithosphere (shaded) descends into the mantle at Pacific subduction zones. They analyse waves that arrive after the direct P waves and before the surface-reflected pP arrivals. 220, 410 and 660, discontinuities at those respective depths.

much better resolution (100 km) and suggest that the 660-km discontinuity deepens slightly more in the immediate vicinity of the subducting slabs. The interesting question, of course, is what changes across these discontinuities, and the measured deflections are much less than predicted for a purely compositional change⁸. So it seems likely that the discontinuity at 660 km depth is due largely to a change of mineral phase.

High-pressure experiments in mineral physics also suggest that both the 410- and 660-km seismic discontinuities are primarily caused by phase changes in olivine resulting from the increased pressure in the upper mantle. The depth at which these phase changes occur depends upon the temperature, so, in principle, observations of variations in the depth can be used to determine *in situ* mantle temperatures. The elevation of the 410-km discontinuity and depression of the 660-km discontinuity are consistent with the effect of a cold subducting slab on the appropriate phase boundaries. However, laboratory values obtained for the scaling factor between pressure and temperature for these reac-

1. Spolsky, C. M., Phillips, C. A. & Uzzell, T. *Nature* **356**, 706-708 (1992).
2. Hedges, S. B., Bogart, J. P. & Maxson, L. R. *Nature* **356**, 708-710 (1992).
3. Quattro, J. M., Avise, J. C. & Vrijenhoek, R. C. *Proc. natn. Acad. Sci. U.S.A.* **89**, 348-352 (1992).
4. Maynard Smith, J. *The Evolution of Sex* (Cambridge University Press, 1978).

tions vary by up to a factor of two, limiting the accuracy with which mantle temperature differences can be inferred.

The character of the 660-km discontinuity bears on the issue of whether the mantle convects as a whole unit or whether the boundary divides the mantle into two separate convection cells. The fate of the subducting oceanic slabs when they reach the discontinuity is one of the clues used to address this ques-

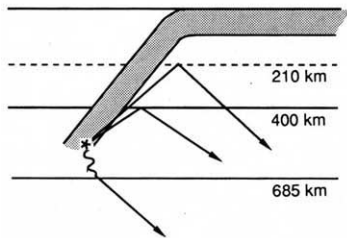


FIG. 2 S waves (wavy line) are converted into P waves on transmission through the (depressed) 660-km boundary. Vidale and Benz also identify P waves reflected from the undersides of less deep interfaces.

tion. Thermal models which show subducting slabs passing cleanly through the 660-km discontinuity into the lower mantle predict greatly reduced mantle temperatures in a narrow zone within the subducting slab itself^{9,10}. These models do not predict the depression seen by Vidale and Benz 100 km to the side of the slab, nor the 1,500-km-wide regional depression seen in our long-period data⁷. The most likely explanation for these anomalies is that the slab broadens or turns near 660 km, thus spreading the cold slab temperatures into the adjacent mantle. This model is supported by recent seismic velocity inversions^{11,12} for northwest Pacific subduction zones which image horizontal extensions of the slabs just above 660 km. The apparent deflection of the slabs could result from resistance to slab penetration through the 660-km phase change, as some

models of mantle convection have indicated^{13,14}. Although these results argue against unhindered slab penetration into the lower mantle, they do not settle the long-standing controversy regarding whether convection is stratified near 660 km or not, as the slab material might eventually make its way into the lower mantle. (For more on this, see the recent News and Views article by Lay¹⁵.)

Among the most intriguing results of Vidale and Benz's analysis are the apparent reflections from a discontinuity near 210 km, which the authors suggest is the base of the asthenosphere, a plastic zone below the lithosphere. For years, seismologists have found sporadic evidence for a 220-km discontinuity in particular

areas, but recent global studies indicate that this is restricted to certain continental regions and is not a significant global feature^{3,16}. Vidale and Benz present a puzzle: if 220-km discontinuities are seen in locations as diverse as subduction zones and continental shields, why are they not observed elsewhere? One possible explanation, suggested by Vidale and Benz, is that the interface is laminated, making it more sensitive to the short-period data used in their analyses than to the long-period data used in the global analyses. □

Peter Shearer is in the Institute of Geophysics and Planetary Physics, University of California, San Diego, California 92093, USA.

RAS REGULATION

NF is enough of GAP

Gideon Bollag and Frank McCormick

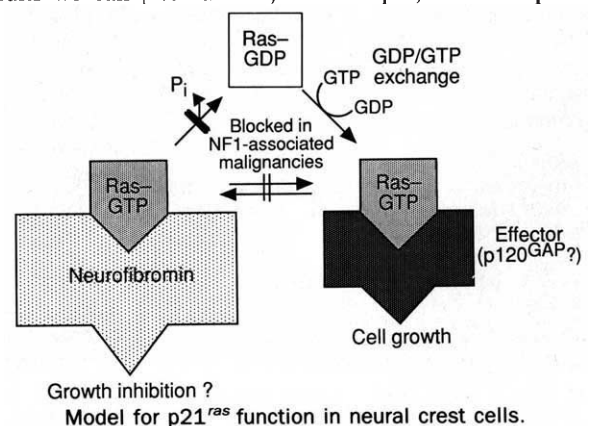
NEUROFIBROMATOSIS type 1 (NF1) is a genetic disease characterized by benign and sometimes malignant tumours that originate in the neural crest. Two years ago, identification of the gene which is defective in NF1 patients revealed a telling pointer to its normal cellular function and its role in the disease: the gene product, now known as neurofibromin, is a GTPase-activating protein (GAP) for the p21^{ras} proto-oncogene product that converts active GTP-bound p21^{ras} to its inactive GDP-bound state¹.

Following up this pointer, Basu *et al.* (on page 713 of this issue²) and DeClue *et al.* (in the latest issue of *Cell*³) now show that cells derived from malignant NF1 tumours express very low levels of neurofibromin. As a result, p21^{ras} accumulates in its GTP-bound state and thus contributes to the transformed state of the cells through interaction with certain effectors (as yet unknown). This is reminiscent of the role of p21^{ras} in other tumour types; oncogenic mutations that render p21^{ras} insensitive to GAP activity result in an accumulation of p21^{ras}-GTP and hence uncontrolled growth. From these new results we can infer that neurofibromin is a tumour suppressor, as its loss of function contributes to tumour growth.

Calculations based on the intrinsic kinetic parameters of purified p21^{ras} proteins suggest that, in the absence of other factors, about 50 per cent of p21^{ras} should be in the GTP state. In normal Schwann cells (as in other normal cells), GTPase activation reduces this level to about 5 per cent. In the NF1-derived malignant

Schwannomas that lack neurofibromin but express normal levels of p120^{GAP}, the p21^{ras}-GTP pool approaches the 50 per cent level indicative of null GAP activity^{2,3}. So it seems that, in Schwann cells, p120^{GAP} has little or no part in p21^{ras}-GTPase activation, and that neurofibromin is the major negative regulator. Conversely, in mouse fibroblasts p120^{GAP} activity is enough to maintain p21^{ras} in its inactive state despite the presence of neurofibromin⁴. These observations imply that although neurofibromin and p120^{GAP} are both ubiquitous, one or the other predominates to determine the fraction of p21^{ras} that is in the active GTP state.

What then is the function of the alternative p21^{ras} regulator (p120^{GAP} in Schwann cells, neurofibromin in fibroblasts)? One possibility is that p120^{GAP} and neurofibromin are involved in signal emission from p21^{ras}-GTP as well as downregulation (see figure)⁵. In this context, it is informative to note that p21^{ras} can send both positive and negative growth signals in cells derived from the neural crest. In PC12 pheochromocytoma cells, for example, activated p21^{ras}



- Vidale, J. E. & Benz, H. M. *Nature* **356**, 678-683 (1992).
- Revenaugh, J. & Jordan, T. H. *J. geophys. Res.* **96**, 19,763-19,780 (1991).
- Shearer, P. M. *J. geophys. Res.* **96**, 18,147-18,182 (1991).
- Barley, B. J., Hudson, J. A. & Douglas, A. *Geophys. J. R. astr. Soc.* **69**, 159-172 (1982).
- Bock, G. & Ha, J. *Geophys. J. R. astr. Soc.* **77**, 593-615 (1984).
- Richards, M. A. & Wicks, C. W. *Geophys. J. Int.* **101**, 1-35 (1990).
- Shearer, P. M. & Masters, T. G. *Nature* **355**, 791-796 (1992).
- Christensen, U. R. & Yuen, D. A. *J. geophys. Res.* **89**, 4389-4402 (1984).
- Schubert, G., Yuen, D. A. & Turcotte, D. L. *Geophys. J. R. astr. Soc.* **42**, 705-735 (1975).
- Creager, K. C. & Jordan, T. H. *J. geophys. Res.* **91**, 3573-3589 (1986).
- van der Hilst, R. *et al.* *Nature* **353**, 37-43 (1991).
- Fukao, Y. *et al.* *J. geophys. Res.* (in the press).
- Machetel, P. & Weber, P. *Nature* **350**, 55-57 (1991).
- Pelteir, W. R. & Solheim, L. P. *Geophys. Res. Lett.* **19**, 321-324 (1992).
- Lay, T. *Nature* **355**, 768-769 (1992).
- Revenaugh, J. & Jordan, T. H. *J. geophys. Res.* **96**, 19,781-19,810 (1991).

causes differentiation. In primary Schwann cells, p21^{ras} proteins alone again inhibit cell growth, yet in collaboration with nuclear oncogenes they stimulate cell proliferation⁶. If neurofibromin mediates the negative growth signal, then loss of neurofibromin could contribute to transformation both through loss of a negative signalling pathway and by allowing p21^{ras}-GTP to accumulate and activate other effectors (p120^{GAP}, for example). This hypothesis is also consistent with the failure to detect mutations to the p21^{ras} gene in neural crest-derived tumours⁷.

The benign tumours characteristic of NF1 consist largely of normal-looking Schwann cells. These cells probably contain a normal neurofibromin gene as well as the defective allele, and are likely to express intermediate levels of p21^{ras}-GTP. In the malignant tumours which sometimes develop, the remarkably low level of neurofibromin is probably due to a 'second hit' in the other neurofibromin allele or in another gene that regulates neurofibromin expression. These second hits could be responsible for progression from benign to malignant tumours in NF1, although such progression has not yet been clearly documented. It will be of considerable interest to determine whether these cells proliferate because they have lost a negative effector of p21^{ras} (as expected from the behaviour of normal Schwann cells) or because the accumulated p21^{ras}-GTP provokes abnormal proliferation (as expected from the behaviour of NF1 Schwannomas). In either case, the identification of negative and positive effectors of p21^{ras} will lead to a better understanding of the involvement of p21^{ras} in neural crest-cell growth and differentiation.

Several implications for the treatment of tumours associated with neurofibromatosis are raised by these considerations. One aim is the inactivation of p21^{ras}, which is already a target for cancer therapeutics. Furthermore, because p120^{GAP} is present but apparently latent, activators of p120^{GAP} should be particularly beneficial. Finally, inhibition of GDP/GTP exchange should also counteract the loss of neurofibromin in these tumours. We expect that progress in understanding NF1 and the p21^{ras} signalling pathway will continue to be complementary. □

Gideon Bollag and Frank McCormick are in the Department of Molecular Biology, Chiron Corporation, Emeryville, California 94608, USA.

Further up the kinetic pathway

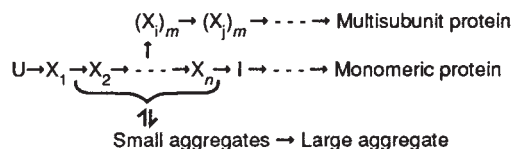
Roger Pain

Of all macromolecules, proteins are the most liable to dangerous liaisons that lead ultimately to death by aggregation. It would seem that families of chaperones have evolved to keep immature proteins on the straight and narrow by protecting them from improper associations. Unlike their victorian prototypes who, by all accounts, operated singly and on a one-to-one basis, it is becoming clear that a number of chaperone-associated factors act in co-operation to assist the folding protein to achieve its functional conformation. On page 683 of this issue¹, members of Ulrich Hartl's laboratory (Langer *et al.*) propose a pathway of co-operative interaction that parallels the protein-folding pathway. They also show that a chaperone protein, GroEL, can cycle to 'catalyse' the transfer of protein from a pool, in which it is sequestered in a partially folded conformation, to the native, folded state.

The evidence that proteins refold *in vitro* along a pathway was reviewed in these columns last month². Several lines of evidence suggest a hierarchical element of folding, with increasing association of hydrophobic groups and increasing compaction following the formation, within milliseconds, of secondary structure³ (see figure). A highly compact intermediate forms late in the folding pathway, within 100 ms or so, differing from the native state chiefly in its lack of persistent tertiary interactions^{4,5}. Modelling studies lead to the conclusion also that the 'unfolded' state of a protein is highly condensed under strongly folding conditions⁶ but there is uncertainty as to the timescale of collapse from the extended state that exists in strong denaturant solution.

A common feature of protein folding is that some intermediates exhibit a considerable tendency to aggregate. This tendency is particularly evident for multisubunit proteins, which *in vitro* have to be refolded at high dilutions. Successful folding *in vitro* and *in vivo* therefore depends on the competition between productive folding and non-specific association (see figure). This has led to intensive studies of heat-shock proteins and their possible role in favouring the pathway towards functional protein and preventing aggregation⁷.

There are equivalents of eukaryotic heat-shock proteins in *Escherichia coli*. GroEL has been shown to bind to a folding intermediate of Rubisco monomer, preventing the strong tendency to aggregate. The addition of GroES in the presence of ATP leads to the release and specific dimerization to functional protein⁸. DnaJ and GrpE are two further heat-shock proteins found in *E. coli* that are known to cooperate with a third, DnaK, in promoting correct folding of proteins, whose variety suggests little specificity in the interaction. Protein folding is one of the most important processes within a living cell, so how



Scheme for the folding and assembly pathway *in vitro* of monomeric and multisubunit proteins. X_i represents folding intermediates that have been detected experimentally by a variety of techniques (see, for example, ref. 3). Secondary structure is present in X_1 . Collapse is evident between X_1 and I, although the degree of compactness has not been well defined until I, which has been well characterized in those cases where it accumulates significantly during folding. Monomers destined for multisubunit proteins associate early in the folding pathway, after which further conformational rearrangement takes place to give the functional protein. Aggregation may take place from a number of 'sticky' intermediates between X_1 and I, the initial but not the later stages being reversible and hence susceptible to chaperone interaction. This would be one possible basis of the 'refolding' activity reported for chaperone proteins.

these different factors interact with the folding protein, how and if they relate to one another, and how the folding protein is released in a form that will not aggregate, are questions of considerable interest.

In tackling some of them, Langer *et al.* used two model proteins, rhodanese and α -lactalbumin. Rhodanese is notoriously difficult to refold *in vitro*, the most successful protocols involving the use of detergents to prevent aggregation⁹. It has been shown to form a compact intermediate when diluted from strong denaturant. When diluted into solutions of molar excess of DnaK and DnaJ, respectively, the aggregation of rhodanese was reduced by binding to each protein in its partially folded state. The addition of Mg-ATP to the complex with DnaK led to release of rhodanese and subsequent aggregation. The interesting finding was that DnaK, together with substoichiometric amounts of DnaJ, bound and stabilized rhodanese during refolding much more effectively than either alone, and this stabilization

1. Xu, G. *et al.* *Cell* **62**, 599-608 (1990).
2. Basu, T. N. *et al.* *Nature* **356**, 713-715 (1992).
3. DeClue, J. E. *et al.* *Cell* **69**, 265-273 (1992).
4. Zhang, K. *et al.* *Science* **254**, 1630-1634 (1991).
5. Bourne, H. R., Sanders, D. A. & McCormick, F. *Nature* **348**, 125-132 (1990).
6. Ridley, A. J. *et al.* *EMBO J.* **7**, 1635-1645 (1988).
7. Bos, J. L. *Cancer Res.* **49**, 4682-4689 (1989).

was greatest in the presence of Mg-ATP. So DnaJ significantly modifies the protective function of DnaK.

In an attempt to mimic conditions under which DnaK and DnaJ might bind to a protein being synthesized on the ribosome, reduced and carboxymethylated α -lactalbumin was used. This has been shown to be unfolded, with no secondary structure. DnaK bound efficiently to this extended protein, in contrast to DnaJ and GroEL which bind preferentially to the more collapsed intermediate form. Interestingly, in the presence of DnaJ, DnaK also binds preferentially to the collapsed form, further underlining the modulation of DnaK by DnaJ. Thus the two heat-shock proteins together exhibit similar behaviour to the chaperone GroEL.

Following what is already known of GroEL/GroES as a chaperone combination that leads to correct folding, Hartl and his colleagues looked to see if the refolding protein could be transferred from its complex with DnaK/DnaJ to GroEL/GroES in the presence of Mg-ATP. No reactivation of rhodanese was obtained. Excitingly, however, the addition of GrpE to the mixture led to rapid and efficient folding to active enzyme. For optimum folding, all the factors were found to be necessary. The go-between role of GrpE strongly supports the proposal that a chaperone pathway exists and is parallel to the protein-folding pathway.

DnaK could bind first to the extended form of the incompletely synthesized protein, and subsequent binding of DnaJ would allow the protein to fold to a more condensed form while still remaining bound and protected against aggregation. Finally, the partially folded protein would be transferred under the influence of GrpE to the GroEL/GroES complex from which it is released in the folded form with the hydrolysis of ATP.

Much ink has been spilt in trying to decide whether the word 'catalysis' should be used to describe chaperone function. The final experiment described by Langer *et al.* shows that the transfer of protein from the DnaK/DnaJ complex can be effected by substoichiometric amounts of GroEL. This suggests that a pool of protein, protected in its partially

folded form by DnaK/DnaJ, can be transferred at an increased rate to its folded form by a cycling process involving GroEL, which therefore comes much nearer to the usually accepted definition of a catalytic process.

Fascinating problems remain — the processes of transfer between chaperone proteins, the mechanism of release and the precise location at which folding occurs are still not fully understood at the molecular level. The present work, carried out with purified components *in vitro*, is several steps away from the living cell. An *E. coli* mutant lacking DnaK grows, albeit slowly. There are other functions known for DnaJ, DnaK

CARBON CYCLE

Catalytic conversions

J. R. Toggweiler

It is not an ordinary day in science when someone claims to discover a new pool of biospheric carbon which is larger than the carbon pools in terrestrial plants or atmospheric CO₂. In 1988, Sugimura and Suzuki described a new method for the high-temperature catalytic oxidation of dissolved organic carbon (DOC) in seawater^{1,2} which, they claimed, extracted up to four times more organic carbon from near-surface oceanic waters and two times more carbon from deep waters than existing wet-chemical oxidation methods. Two papers in this issue of *Nature*^{3,4} and one recently in *Science*⁵ present some of the first published re-examinations of these results. Although contradictory in many respects, the reports certainly take some of the lustre off the original results.

The oceanic DOC pool revealed by the older wet methods is large enough by itself (700×10⁹ tons C) to rival the terrestrial biomass or atmospheric CO₂ reservoirs. The conventional wet-chemistry methods oxidize dissolved organic carbon in seawater with either persulphuric acid or high-intensity ultraviolet light. By burning DOC on a hot platinum catalyst, Sugimura and Suzuki claimed to extract enough additional carbon to add 1,000×10⁹ tons C or so to the oceanic DOC pool. While the 'old' DOC is known to be thousands of years old⁶ and its variation within the ocean is rather minimal, the so-called new DOC appears to have high surface concentrations and large gradients within the ocean's main thermocline. These qualities suggest that the new DOC must be produced and broken down by open-ocean plankton communities within a decade or two. So quick a throughput of such a large amount of carbon would make the new DOC one of the most active carbon reservoirs on the planet.

and GrpE, for example in bacteriophage assembly, and, for the mitochondrial homologues, in membrane translocation. But proteins have a much more dynamic lifestyle than was once thought to be the case and a chaperone pathway, such as that persuasively argued for by Langer *et al.*, could constitute a common mechanism to protect proteins as they go about their business of folding, unfolding, associating and disassociating in the living cell. □

Roger Pain is in the Department of Biochemistry and Genetics, The Medical School, University of Newcastle upon Tyne, Newcastle upon Tyne NE2 4HH, UK.

Martin and Fitzwater, on page 699 of this issue³, use the catalytic method to analyse vertical profiles of seawater from 59.5° N in the North Atlantic, from Drake Passage and from the equatorial Pacific. They find very little change between stations in DOC concentrations in deep water in spite of very large differences in ¹⁴C age and in dissolved oxygen levels at the three locations. This result suggests that bacterial consumption of DOC in the ocean's deep interior does not account for much of the observed consumption of oxygen. Martin and Fitzwater do, however, find a very strong DOC gradient in surface waters between the equator (130 μ M) and 9° N (230 μ M) in the equatorial Pacific. The DOC level in high-salinity equatorial water, which has a Southern Hemisphere source⁷, is much lower than the DOC level in the low-salinity water of the north equatorial countercurrent.

Martin and Fitzwater's 100- μ M DOC decrease between 9° N and the equator is similar to the vertical DOC decrease in the upper few hundred metres of the North Pacific reported by Sugimura and Suzuki. Overall, Martin and Fitzwater report that their catalytic apparatus produces excellent agreement with Suzuki's when they analyse the same water. They support the assertion that the oceans contain about 1,000×10⁹ tons of a previously unknown pool of DOC.

According to Sugimura and Suzuki, the new DOC is biologically labile, yet resistant to attack by the strong chemical oxidants employed by the wet chemical methods. Ogawa and Ogura, on page 696 of this issue⁴, revisit this question by repeating the comparison between the two methods. Ogawa and Ogura find that their catalytic apparatus extracts only about one-third more carbon overall than does wet chemistry. Their sam-

1. Langer, T., Lu, C., Echols, H., Flanagan, J., Hayer, M. K. & Hartl, F. U. *Nature* **356**, 683–689 (1992).
2. Creighton, T. E. *Nature* **356**, 194–195 (1992).
3. Christensen, H. & Pain, R. H. *Eur. Biophys. J.* **19**, 221–229 (1991).
4. Pititsyn, O. B., Pain, R. H., Semisotnov, G. V., Zerovnik, E. & Razgulyaev, O. I. *FEBS Lett.* **262**, 20–24 (1990).
5. Roder, H., Elve, G. A. & Englander, S. W. *Nature* **335**, 700–704 (1988).
6. Shortle, D., Chan, H. & Dill, K. A. *Protein Sci.* **1**, 201–205 (1992).
7. Ellis, R. J. *Nature* **328**, 378–379 (1987).
8. Goloubinoff, P., Christeller, J. T., Gatenby, A. A. & Lorimer, G. H. *Nature* **337**, 44–47 (1989).
9. Tandon, S. & Horowitz, P. M. *J. biol. Chem.* **265**, 5967–5970 (1990).

ples come from near-shore environments south of Tokyo, so that they cannot compare the variations between the two techniques across the oceanographic DOC gradients reported by Sugimura and Suzuki and by Martin and Fitzwater.

High-mass fractions

Sugimura and Suzuki separated DOC into fractions of different relative molecular mass (M_r) using a gel-filtration technique in which low- M_r material is preferentially retained during the passage of a DOC mixture through a column. Sugimura and Suzuki claimed that much of the new DOC is high-molecular-mass material ($M_r > 20,000$). Ogawa and Ogura concentrate two high- M_r fractions by forcing water under pressure through nanometre-sized pores (ultrafiltration). They find very little DOC with molecular masses above 10,000. Using each method they recover the same amount of carbon, both with $M_r > 1,000$ and with $M_r > 10,000$. The authors suggest that there might be no chemically refractory, high- M_r DOC at all in seawater.

Benner *et al.*, in *Science*⁵, attempt to characterize DOC by concentrating the $M_r > 1,000$ fraction using the same ultrafiltration technique as Ogawa and Ogura. They find that a third of the surface-water DOC falls in this size fraction, a result which agrees very well with Ogawa and Ogura's. This concentrate also contains much less nitrogen (C/N = 15) than the high- M_r fractions extracted by Sugimura and Suzuki^{1,8} (C/N = 7). Benner *et al.* find that high- M_r polysaccharides make up about half of the $M_r > 1,000$ fraction in surface water, whereas aromatic or olefinic carbons make up a high proportion of that fraction in deep water. Although they disagree with many of Sugimura and Suzuki's results, Benner *et al.* "concur that the high- M_r components in the upper ocean are reactive and support much of the heterotrophic activity in the surface ocean".

In all the new results, only two direct comparisons can be made. Both Martin and Fitzwater³ and Benner *et al.*⁵ measured water from the Hawaii Ocean Time Series station which was collected for the Joint Global Ocean Flux Study (JGOFS) DOC Workshop (Seattle, July 1991)⁹. Martin and Fitzwater report 152 μM for surface water, whereas Benner *et al.* report only 82 μM . The same level of disagreement exists for water at the O_2 minimum (86 compared with 38 μM) and for deep water (112 compared with 41 μM). Benner *et al.* subtract a system blank of 22 μM from their raw data, whereas Martin and Fitzwater do not subtract any such blank. It is noteworthy, perhaps, that the difference between Martin and Fitzwater and Benner

et al. is almost the same for both the surface and deep samples (70 μM).

The subject of system blanks is probably at the heart of many of the disagreements arising from the catalytic method⁹. When ultrapure distilled water is run through the high-temperature catalyst, significant DOC levels can result. Suzuki *et al.*¹⁰ report values of 15–30 μM ; Ogura and Ogawa, 10 μM . Catalytic oxidation always recovers more carbon from distilled water than does wet oxidation. Suzuki *et al.* and Ogura and Ogawa assume that the carbon in their distilled-water blanks comes from the distilled water, and do not subtract any blank from their seawater measurements. Benner and Strom¹¹ conclude that the carbon in distilled-water blanks is actually carbon released from the catalyst itself. They observe blank values ranging from 10 to 50 μM from different catalysts. Benner *et al.*⁵ treat their 22 μM blank as a system blank and subtract it from their seawater determinations.

High turnover

Is there a 'new' DOC pool, or is it an artefact of the measurements? It remains to be seen whether Sugimura and Suzuki¹ and Martin and Fitzwater's³ high DOC levels suffer contamination by blank carbon. If their high DOC numbers are ultimately shifted down, the size of the global new DOC pool will shrink accordingly. In the end, however, the size of the pool is less interesting than its activity and its role in the marine carbon cycle and in marine ecosystems. The most intriguing aspects of Sugimura and Suzuki's and of Martin and Fitzwater's data are the large vertical and horizontal DOC gradients they reveal in the upper ocean. These gradients cannot be maintained unless large amounts of DOC production and remineralization are taking place. There is no reason at present to think that a blank correction, however large, will eliminate these gradients. The most pressing need is to verify that these gradients are real, and if so, to see how much of the variation is identified by wet oxidation. □

J. R. Toggweiler is in the Geophysical Fluid Dynamics Laboratory, National Oceanic and Atmospheric Administration, Princeton University, PO Box 308, Princeton, New Jersey 08542, USA.

1. Sugimura, Y. & Suzuki, Y. *Mar. Chem.* **24**, 105–131 (1988).
2. Toggweiler, J. R. *Nature* **334**, 468 (1988).
3. Martin, J. H. & Fitzwater, S. E. *Nature* **356**, 699–700 (1992).
4. Ogawa, H. & Ogura, N. *Nature* **356**, 696–699 (1992).
5. Benner, R. *et al. Science* **255**, 1561–1564 (1992).
6. Williams, P. M. & Druffel, E. R. M. *Nature* **330**, 246 (1987).
7. Tsuchiya, M. *et al. Prog. Oceanogr.* **23**, 101–147 (1990).
8. Suzuki, Y. *et al. Mar. Chem.* **16**, 83–97 (1985).
9. Williams, P. M. *US JGOFS News* **3**, 1 (1991).
10. Suzuki, Y. *et al. Deep-Sea Res.* **39**, 185–198 (1992).
11. Benner, R. & Strom, M. *Mar. Chem.* (in the press).

Not making waves

SOME while ago Daedalus devised a 'slippery ship'. Its port and starboard surfaces were electrodes. A current passing between them covered them with a layer of tiny electrolytic gas bubbles, reducing skin-friction almost to zero. In a final stroke of ingenuity, Daedalus propelled his ship electromagnetically. Magnets in the hull surface interacted with the current-carrying water, forcing it backwards and therefore thrusting the ship forwards.

Electromagnetic propulsion has since been taken up in other quarters, notably Japan. So Daedalus is maintaining DREADCO's lead by a further cunning development. A properly designed electromagnetic boat, he claims, need not even make any waves. The last major source of drag can thus be eliminated.

As it advances, a conventional ship pushes water out of the way to make room for its own volume. The water is forced upwards; the resulting waves run endlessly away from the ship, taking energy with them. A submarine has no such losses. The water that it displaces cannot go upwards — the volume that it might occupy is already full. Instead it goes backwards. It flows round the hull at high speed, and decelerates again behind it. The process radiates no energy.

So DREADCO's hydrodynamicists are optimizing the field-pattern of their electromagnetic ship. They are working out the distribution of current density and magnetic-field intensity that will accelerate every element of the water near the ship in the precise pattern expected for the equivalent submarine hull. The mathematics are desperately hard: but success will be wonderful. For the resulting vessel will travel at any speed without making waves. Instead, it will thrust the level water directly backwards around it.

This form of propulsion is utterly efficient. It puts power into accelerating the water around the bow of the vessel, and recovers it again by regenerative electromagnetic braking at the stern. The water is left calm and undisturbed; there is no wake. The vessel will lose no energy to turbulence, and only a little to viscous shear, for its field system imposes perfect laminar flow on the water at every point. Silent and ghost-like, it will seem to slip through the water without even parting it, and will reach vast speeds on almost no power.

The waveless ship will be rapidly adopted by ferry and sea-freight companies. Its furtiveness will also appeal to the world's navies. With no propeller noise, and not disturbing the sea for miles around, it will be that much harder to spot from submarines, aircraft or satellites.

David Jones

Shipwrecked tube worms

SIR — Large vestimentiferan tube worms (Pogonophora, Obturata) are typical members of both hydrothermal and cold seep communities in the Pacific Ocean (see, for example, refs 1, 2). In the Atlantic Ocean, Vestimentifera are known only from the western side, at sulphidic hydrocarbon seeps in the Gulf of Mexico at depths of 350–3,266 m (refs 3, 4), at 500 m depth off Guyana⁵ and at 300 m depth off Montevideo⁶. None was observed at hydrothermal vents on the mid-Atlantic Ridge⁷. In September 1991, during initial exploration of the wreck of the cargo ship *Francois Vieljeux*, we recovered vestimentiferan tube worms and bivalve molluscs, *Idasola* sp (K. W. Ockelmann, personal communication). Both animals are known to contain chemoautotrophic symbiotic bacteria. This is the first report of giant tube worms from the east Atlantic and the first time they have been recorded in a habitat other than a vent or seep site.

The wreck lies at 42° 7.95' N 9° 29.69' W, approximately 30 miles west of Vigo, Spain. It foundered during a storm in 1979, after developing a list due to shifting cargo. Containers on the deck were smashed open and the hatch covers



Anterior end of tube of a vestimentiferan. Scale bar, 10 mm. (D. Nicholson.)

on some of the holds were broken before the vessel landed on the sea floor at 1,160 m water depth. Tins of pineapple were scattered over the sea bed and some of them fell into the open holds. When loaded, the upper part of one hold had contained sacks of beans and sunflower seeds on top of polythene-wrapped bales which contained rolls of sisal twine. Some of this was recovered using an orange-peel grab equipped with a video camera. Most of the seeds were washed away during recovery but

some were retained by the bales.

Two vestimentiferan tube worms (see figure) were found lying on top of one of the bales. One of the tubes was 960 mm long with an intact anterior end of 10 mm outside diameter, tapering to 4.5 mm diameter at the broken end. The upper 250 mm had marked growth collars. The soft tissues were not preserved and it is therefore not possible to identify the species. (We extracted high-molecular-mass DNA from the remains of the animal tissue in the tube 5 weeks after recovery but the vestimentiferan material had been lost to dominant fungal DNA (P. W. H. Holland, personal communication).) The second tube was a middle section 1,270 mm long tapering from 6.0 mm diameter to 4.0 mm. On recovery the upper 200–240 cm of the tubes was a pale straw colour. The lower black portion of the tubes is believed to have been buried in the rotting beans which were black and pulpy when recovered. The tubes are similar in appearance to those of *Lamellibrachia barhami* (Webb)^{1,8} found in water at depths of 1,125–2,050 m off the coasts of Oregon and California, but have smaller collars than two *Lamellibrachia* species from the Atlantic coast of South America^{5,6}.

Bacterial breakdown of the beans would provide substrates for sulphate-reducing bacteria, creating a hydrogen sulphide-rich habitat for the Vestimentifera. Assuming that a minimum of 2 years was required for initial breakdown and colonization by vestimentiferan larvae from presumed nearby cold-seep habitats on the slope, the animals must have grown to their present size in 10 years or less. This gives minimum growth rates of 9.5 and 12.7 cm per year, which compares with observed rates of 9.2–50 cm per year for Vestimentifera from warmer hydrothermal vent habitats². At the depth where the wreck lies there is a strong northgoing flow of warm, high-salinity water of Mediterranean origin (10.7–10.9 °C and 36.12–36.17 ‰)^{9,10}. For vestimentiferans to colonize the wreck under these conditions suggests the existence of nearby sources of larvae.

The animals in chemosynthesis-based communities that obtain nutrition from symbiotic chemoautotrophic bacteria are clearly opportunist in colonizing habitats that contain sufficient sulphide. Such sites include hydrocarbon seeps, sunken wood, whale skeletons¹¹ and, as we now report, organic-rich ship cargoes. The steep continental slope of the north-eastern Atlantic is difficult to explore with surface vessels and has been subject to few observations from submersibles. Our discovery considerably extends the

known range of large tube worms and points to the occurrence of cold saline or hydrocarbon seeps along the Atlantic slope of the Iberian peninsula, providing a normal habitat for these animals. Vestimentifera may be present along most continental margins at seep sites.

P. R. DANDO

A. J. SOUTHWARD

E. C. SOUTHWARD

Marine Biological Association of the United Kingdom,

D. R. DIXON

Plymouth Marine Laboratory, Citadel Hill, Plymouth PL1 2PB, UK

A. CRAWFORD

M. CRAWFORD

Deep Water Recovery and Exploration Ltd, Kilburns House, Newport-on-Tay, Fife, DD6 8PL, UK

1. Jones, M. L. *Bull. Biol. Soc. Wash.* **6**, 117–158 (1985).
2. Tunnicliffe, V. A. *Rev. Oceanogr. Mar. Biol.* **29**, 319–407 (1991).
3. Paull, C. K. *et al. Science* **226**, 965–967 (1984).
4. MacDonald, I. R. *et al. Geo-mar. Lett.* **10**, 244–252 (1990).
5. Land, J. van der & Nørrevang, A. *Kong. Dansk. vidensk. Selskab. Biol. Skrift.* **21**(3), 1–102.
6. Mañé-Garzon, F. & Montero, R. *Rev. Biol. Uruguay* **8**, 1–28 (1985).
7. Karson, J. A. & Brown, J. R. *Mar. geophys. Res.* **10**, 91–107; 109–138 (1988).
8. Webb, M. *Bull. mar. Sci.* **19**, 18–47 (1969).
9. Lacombe, H. & Tchermia, P. *Cah. Oceanogr.* **12**, 417–439 (1960).
10. Madelaine, F. *Cah. Oceanogr.* **19**, 125–136 (1967).
11. Smith, C. R. *et al. Nature* **341**, 27–28 (1989).

Cross-pollination

SIR — Questions raised by Madsen *et al.*¹ and by Parker² on the adaptive value of multiple mating for female adders are remarkably similar to those being asked by plant reproductive ecologists. Multiple pollinations of a flower increase the size and genetic diversity of the pollen load^{3,4}, which in turn may affect the intensity of 'pollen competition' and the opportunity for female 'mate choice'^{5–7}. It is often possible to determine *in vivo* performance of pollen (for example, rate of pollen-tube elongation), and performance can vary with the intensity of pollen competition as well as with male identity^{8,9}. Parker finds it improbable that mate choice could evolve on the basis of parental combination, but this is one way to describe self- and cross-incompatibility reactions in higher plants, and pollen performance may vary with parental combination more generally¹⁰.

The link between pollen performance and progeny fitness is under investigation. A substantial fraction of the genes expressed by haploid pollen also is expressed in diploid progeny^{11,12}. Further, large pollen loads sometimes are associated with reduced fruit abortion and more vigorous progeny^{13,14}, and parental combinations that yield better pollen performance may also produce the most

seeds¹⁵. Substituting 'sperm' for 'pollen', one immediately sees the overlap with issues discussed in the adder example. Botanists and zoologists would benefit from more cross-pollination concerning these issues.

NICKOLAS M. WASER
MARY V. PRICE

Department of Biology,
University of California,
Riverside, California 92521, USA

ROBERT I. BERTIN

Department of Biology,
College of the Holy Cross,
Worcester, Massachusetts 01610, USA

1. Madsen, T., Shine, R., Loman, J. & Håkansson, T. *Nature* **355**, 440–441 (1992).
2. Parker, G. A. *Nature* **355**, 395–396 (1992).
3. Marshall, D. L. & Ellstrand, N. C. *Am. Nat.* **127**, 446–461 (1986).
4. Broyles, S. B. & Wyatt, R. *Evolution* **44**, 1454 (1990).
5. Willson, M. F. & Burley, N. *Mate Choice in Plants: Tactics, Mechanisms, and Consequences* (Princeton University Press, 1983).
6. Lyons, E. E., Waser, N. M., Price, M. V., Antonovics, J. & Motten, A. F. *Am. Nat.* **134**, 409–433 (1989).
7. Marshall, D. L. & Folsom, M. W. *Am. Rev. Ecol. Syst.* **22**, 37–63 (1991).
8. Ter-Avanesian, D. V. *Theor. appl. Genet.* **52**, 77 (1978).
9. Snow, A. A. & Spira, T. P. *Nature* **352**, 796 (1991).
10. Waser, N. M., Price, M. V., Montalvo, A. M. & Gray, R. N. *Evol. Trends Plants* **1**, 29–33 (1987).
11. Mulcahy, D. L. *Science* **206**, 20–23 (1979).
12. Tanksley, S. D., Zamir, D. & Rick, C. M. *Science* **213**, 453–455 (1981).
13. McKenna, M. A. in *Biotechnology and Ecology of Pollen* (ed. D. L. Mulcahy) 443 (Springer, New York, 1986).
14. Stephenson, A. G., Windsor, J. A. & Davis, L. E. in *Biotechnology and Ecology of Pollen* (ed. D. L. Mulcahy) 429–434 (Springer, New York, 1986).
15. Waser, N. M. & Price, M. V. *Ecology* **72**, 171 (1991).

Angiogenic factor

SIR — Platelet-derived endothelial cell growth factor (PD-ECGF), a protein of relative molecular mass 45,000¹, stimulates endothelial cell growth and chemotaxis *in vitro* and angiogenesis *in vivo*. It lacks a hydrophobic signal sequence and is not a classical secretory protein². We have isolated cDNA clones for human thymidine phosphorylase (dThdPase) which catalyses the reversible phosphorylation of thymidine, deoxyuridine and their analogues to their respective bases and 2-deoxyribose 1-phosphate, and is essential to nucleic acid metabolism because it regulates the availability of thymidine. We find that 120 amino acids of human dThdPase are identical to the sequence of PD-ECGF.

We have designed primers for the polymerase chain reaction (PCR) based on the peptide sequence data derived from *Achromobacter lyticus* protease I (lysyl endopeptidase) fragments of purified human dThdPase³. PCR products from amplification of the human liver cDNA were subcloned and sequenced. Amino-acid sequences were deduced from nucleotide sequences of the longest clone (288 base pairs) and compared with sequences in the NBRF amino-acid databank. This sequence was 100% identical to the sequence of PD-ECGF (re-

sidues 149–244). The amino acid sequence of all four lysyl endopeptidase fragments of dThdPase could be aligned with the amino acid sequence of PD-ECGF (residues 125–139, 140–157, 158–178 and 236–244, respectively). Our data indicate that residues 125–244 of PD-ECGF are identical to the sequence of dThdPase. The amino-acid sequence of the human dThdPase clone is 61.9% homologous to the sequence of *Escherichia coli* dThdPase. The total amino-acid sequence of *E. coli* dThdPase⁴ has 39.3% homology with PD-ECGF². Amino-acid sequences of the putative thymidine binding site on *E. coli* dThdPase⁴ are very similar to the sequence of PD-ECGF. Ishikawa *et al.*² reported that there is only one copy of the PD-ECGF gene in the human genome. PD-ECGF and human dThdPase thus seem to be derived from the same gene, which should help to elucidate how PD-ECGF works as an angiogenesis factor.

TATSUHIKO FURUKAWA
AKIHIKO YOSHIMURA
TOMOYUKI SUMIZAWA
MISAKO HARAGUCHI
SHIN-ICHI AKIYAMA

Institute of Cancer Research,
Kagoshima University,
8-35-1 Sakuragaoka,
Kagoshima 890, Japan

KIYOSHI FUKUI

UBE Laboratory, UBE Industries Ltd,
Yamaguchi 755, Japan

MINORU ISHIZAWA

UBE Bioscience Laboratory,
Research Institute of Biosciences,
Science University of Tokyo,
Chiba 278, Japan

YUJI YAMADA

Taiho Pharmaceutical Co Ltd,
Saitama 357, Japan

1. Miyazono, K., *et al.* *J. biol. Chem.* **262**, 4098 (1987).
2. Ishikawa, F. *et al.* *Nature* **338**, 557–562 (1989).
3. Yoshimura, A. *et al.* *Biochim. biophys. Acta* **1034**, 107–113 (1990).
4. Walter, M. R. *et al.* *J. biol. Chem.* **265**, 14016 (1990).

Pulsar planets

SIR — As recently pointed out¹, planets are expected to form around a millisecond pulsar such as PSR1257+12 (ref. 2) if the pulsar originally had a binary companion that was tidally disrupted and formed a disk around the pulsar. The planets could then have condensed out of the disk in ways similar to that which took place in the solar nebula¹. But it has been suggested¹ that the disruption of the (non-degenerate) companion of the pulsar was triggered by the evaporative heating and envelope expansion produced by the irradiation by the pulsar. Here I point out that another way for obtaining a millisecond pulsar surrounded by a massive disk is provided by our white dwarf plus neutron-star merger model for the formation of single

millisecond pulsars³.

Close binaries consisting of a massive ($\geq 0.7 M_{\odot}$) white dwarf and a neutron star do exist: PSR0655+64 is an example. Although the orbital period of this system (about 24 h) is too long to cause it to merge (due to gravitational radiation losses) within a Hubble time, such systems are also expected to be formed with shorter orbital period⁴. They are the end products, after a phase of common envelope evolution, of B-emission X-ray binaries in which the Be star had a mass $\leq 6-8 M_{\odot}$ (refs 4,5). When a white dwarf plus neutron-star binary has an orbital period ≤ 15 h, it will merge by gravitational radiation losses within a Hubble time.

This is expected to happen to a considerable fraction of such systems^{4,5}. When the orbital period has decreased so far that the white dwarf begins to overflow its Roche lobe, the resulting mass transfer is, for white dwarf masses $\geq 0.7 M_{\odot}$ (and for an adopted neutron star mass of $1.4 M_{\odot}$), unstable and will cause the white dwarf to be completely disrupted on its dynamical timescale (seconds) and to form a disk around the neutron star that contains the orbital angular momentum of the binary, which is about $6 \times 10^{50} \text{ g cm}^2 \text{ s}^{-1}$. (The instability of the mass transfer is due to the steepness of the mass-radius relation of massive white dwarfs³.) The neutron star will in most cases already be older than about 10^8 yr when this happens, and its magnetic field will have partly decayed⁴. The accretion of a small fraction of the white dwarf matter (about $0.05 M_{\odot}$) is then sufficient to spin it up to a period of the order of milliseconds. The result will be a millisecond pulsar surrounded by a massive disk of relatively heavy elements (C, O, Ne, Mg and heavier).

Planet formation in such a disk is then expected to proceed along lines similar to those outlined in ref. 1. The disk contains sufficient angular momentum and heavy elements (Si and heavier) to form several planets of Earth-like composition, having orbital radii and masses similar to those observed in PSR-1257+12 (ref. 2).

E. P. J. VAN DEN HEUVEL

Center for High Energy Astrophysics
and Astronomical Institute "Anton
Pannekoek",
University of Amsterdam,
1098 SJ Amsterdam, The Netherlands

1. Stevens, I. R., Rees, M. J. & Podsiadlowski, P. *Mon. Not. R. astr. Soc.* **254**, 19P–22P (1992).
2. Wolszczan, A. & Frail, D. *Nature* **355**, 145–147 (1992).
3. Bonsema, P. J. F. & van den Heuvel, E. P. J. *Astr. Astrophys.* **146**, L3 (1985).
4. Bhattacharya, D. & van den Heuvel, E. P. J. *Phys. rep.* **203**, 1–124 (1991).
5. van den Heuvel, E. P. J. in: *X-ray Binaries and Recycled Pulsars* (eds. van den Heuvel, E. P. J. & Rappaport, S. A.) (Kluwer, Dordrecht, in the press).
6. van den Heuvel, E. P. J. & Taam, R. E. *Nature* **309**, 235 (1984).

The new nuclear agenda

William Walker

Stopping the Spread of Nuclear Weapons: The Past and Prospects. By David Fischer. Routledge: 1992. Pp. 336. £40.

Plutonium and Security: The Military Aspects of the Plutonium Economy. Edited by Frank Barnaby. Macmillan: 1992. Pp. 296. £47.50.

FROM the mid-1950s to the late 1980s, attempts to regulate the development and deployment of nuclear weaponry occurred in two largely distinct international frameworks. The first, nuclear arms control, involved the negotiation of a series of bilateral treaties between the dominant nuclear powers, the United States and the Soviet Union. The primary function of the control was to bring

practised over the past three decades has become anachronistic. The great challenge now is to extend existing non-proliferation policy so that it becomes a truly universal framework for disarmament. Furthermore, managing the processes of disarmament, whether in relation to Iraq, North Korea and the remaining threshold countries, or the Ukraine, Kazakhstan and Belarus, or

Russia, the United States and the other states in possession of nuclear weapons, is the most important task now facing the international community.

It is David Fischer's misfortune that this fundamental change in perspective happened after his book was completed in late 1990. The book is nevertheless one that every student of nuclear affairs should acquire. It has three particular merits. The first is the quality of its historical narrative. Fischer's long and distinguished career at the International Atomic Energy Agency (IAEA), and his subsequent role as everyone's favourite adviser, means that he has been in the thick of most of the important developments in nuclear relations since the 1950s. The book's second

virtue is Fischer's detailed examination of the various facets of the nonproliferation regime (or system as he prefers to call it), and especially of the safeguards system on which he is the leading authority. Finally, the book is scrupulously balanced and fair. Although some will find the author too kind to the nuclear industry, there is little of the partisanship and special pleading that mar so many other books on the subject.

My one disappointment is that Fischer says so little about the great personalities in this field — the great facilitators, as well as the awkward customers such as S. Eklund and B. Goldschmidt. Fischer has known them all and has a fund of anecdotes. Besides the entertainment of learning more about them, there is an important question that remains un-

answered in his, as in other, historical accounts: how much has personal influence and prejudice affected the course of nuclear relations? We must wish Fischer a long life so that he will one day feel free to set down on paper his memories of these colourful individuals.

Plutonium and Security is a more specialized book, but in some respects it is more prescient. Its focus is plutonium, which, along with highly enriched uranium, is the material used in the fission stage of nuclear weapons. Chapters by D. Albright and by F. Berkhout cover an important problem that will soon be upon us — the huge flows of plutonium that will result from the operation of the new reprocessing plants at Sellafield in Britain and La Hague in France. Over the past 30 years, around 120 tonnes of plutonium have been separated from civil spent fuels around the world. During the next ten years, 190 tonnes will be separated from these plants alone. Although not covered in this book, this problem will be made still more acute by the extraction of further large quantities of plutonium (perhaps as much as 150 tonnes) from dismantled warheads in Russia and the United States.

The nuclear industry is working hard to develop ways of using plutonium as a reactor fuel. So far, the effort is proving to be an uphill struggle; plutonium fuels are likely to remain much more expensive than uranium fuels, and agencies and governments that engage in plutonium recycling court political unpopularity. The consequence is that much of the plutonium becoming available will have to be dealt with as a waste, rather than as the asset that it has always been assumed to be.

The rest of the book is a mixed bag. Three chapters stand out. One is by Warren Donnelly on the history of US plutonium policy — the retreat from reprocessing and plutonium economies. Another is the excellent discussion by James Lovett of the effort and ingenuity that has gone into developing techniques for safeguarding large reprocessing plants. The third is Gordon Thompson's examination of proposals for halting production of weapon-grade plutonium and uranium in the nuclear-weapon states. This cut-off is today almost a reality; it is to be hoped that the remaining production of plutonium in Russia, and of highly enriched uranium in France, will end soon.

Both of these books are worthwhile contributions, although to read them is to realize how radically the nuclear agenda has changed in the past few months. □

William Walker is in the Science Policy Research Unit, University of Sussex, Mantell Building, Falmer, Brighton BN1 9PF, UK.

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Atomic disposal — soldiers load a warhead into a container on a military truck in the Ukraine in an effort to make the Ukraine a nuclear-free zone (January 1992).

greater stability and predictability to the nuclear relationship between East and West, while allowing either side to continue modernizing and expanding its nuclear forces. The second framework was that provided by a nuclear nonproliferation policy, involving in particular the Euratom Treaty and the Nuclear Non-Proliferation Treaty (NPT), their associated safeguards systems, and the trade regulations set out in the NPT and the Nuclear Suppliers' Guidelines. The main function of the policy was to maintain a condition of perpetual abstinence among countries that did not already possess nuclear weapons.

With the break-up of the Soviet Union and the huge nuclear arms reductions that are following in its wake, this dualism is breaking down. Arms control as

Fire alarm?

Jonathan T. Overpeck

Global Biomass Burning: Atmospheric, Climatic, and Biospheric Implications. Edited by Joel S. Levine. MIT Press: 1991. Pp. 569. \$75, £67.50.

In the arena of global climate change, biomass burning has come to be recognized, along with the burning of fossil fuels and deforestation, as an important source of atmospheric trace gases. As much as 2–5 per cent of the Earth's land area burns each year, mostly as a result of fires caused by humans. This biomass burning causes a net transfer of CO₂, CO, CH₄, NO, NO₂ and N₂O into the atmosphere, generating 5–25 per cent of the total radiative climate forcing due to all anthropogenic greenhouse gas emissions. Locally, biomass burning also contributes to smog, acid precipitation, cloud condensation nuclei and daytime cooling. Although most burning occurs at low latitudes, where it is used for land clearance, agriculture and energy, high-latitude biomass burning also has large environmental effects. Global biomass burning is increasing, making it a global environmental issue that both scientists and policymakers cannot ignore.

This book is a product of the 1990 Chapman conference on global biomass burning and will be a reliable source book for anyone interested in global warming, deforestation, atmospheric chemistry and future climate change. The 63 papers by over 150 contributors include excellent scientific summaries and discussions of new results. Good coverage is given to the detection and estimation of the size of burn area, to the effects of burning on atmospheric chemistry, and to the potential effects of these chemical changes on the climate system. Policy implications are also discussed. The contributor list is inter-

national and the geographical coverage global. Forty pages of bibliography provide a further reason to obtain a copy.

Several general points become clear. First, a great deal of uncertainty is associated with most estimates of biomass burning and rates of associated trace-gas emissions. The study of biomass burning is young, but it overlaps with studies of many other aspects of global change in that they all need a more detailed description of the Earth's surface, including the biosphere. Fortunately, this description is a main aim of the International Geosphere-Biosphere Program (IGBP). Second, the book leaves one with the impression that, despite scientific uncertainty, the problems posed by biomass burning are real and deserve international attention. Policy goals can be set immediately to improve the way in which biomass burning is managed, without waiting for a more complete scientific understanding. This point is true for most global environmental issues, but is too often lost in the unavoidable haze of scientific uncertainty.

Not all of the possible environmental effects associated with biomass burning are covered in the book. As several of the authors make clear, marginally

controlled or uncontrolled fires occur frequently at all latitudes. A prescribed burn can escape its intended perimeter and encroach upon surrounding areas. National parks and nature reserves, perhaps containing the last remnants of particular ecosystems, are thus at risk as long as biomass burning is poorly managed. The effect of climate change on the incidence of fires must also be considered. How will the incidence of "fire weather" alter with changing global climate, and how will this affect efforts to manage biomass burning and protected ecosystems? The book does not answer these questions, but it provides an impetus for further investigations.

In addition to being an important reference book, *Global Biomass Burning* covers ground that is relevant to a wide spectrum of scientists and policymakers concerned with global change. By combining state-of-the-art summaries with new results and quick printing, it is an excellent example of what a conference or workshop book can and should be. □

Jonathan T. Overpeck is with the National Oceanic and Atmospheric Administration Paleoclimatology Program, Geophysical Data Center, 325 Broadway, Boulder, Colorado 80303-3328, USA.

Opening the mind's manacles

John Hardy

Molecular Genetic Approaches to Neuropsychiatric Diseases. Edited by Jurgen Brosius and Robert T. Fremeau. Academic: 1991. Pp. 493. \$69.95, £47.

THE title of this book is misleading. I picked it up expecting to read about the entertaining saga of the applications of genetic linkage analysis to schizophrenia, Alzheimer's disease and affective disorder. But schizophrenia is dealt with in merely a single chapter (and not by any of the groups working on its molecular genetics); the coverage of the molecular biology of Alzheimer's disease, although competent, is way out of date (preceding the identification of pathogenic mutations in the β -amyloid precursor protein); and affective disorder is not mentioned. The book is, in fact, a series of chapters discussing the application of molecular genetics and molecular biology to a variety of diseases, from retinoblastoma to Duchenne's muscular dystrophy, and from Tay-Sachs to schizophrenia. I would guess that the volume will be most useful for neurologists. As such, it is interesting, but suffers from the serious drawback that it has been hopelessly overtaken by events.

Thus, not only is the Alzheimer's story completely outdated, but chapters on neurofibromatosis type 1, myotonic dystrophy and fragile X syndrome all pre-date the cloning of the pathogenic loci. Furthermore, spinal bulbar muscular atrophy, and Charcot-Marie-Tooth and prion diseases receive no coverage, despite the fact that their underlying molecular genetics is now almost completely known.

Chapters on clone crunching by members of H. Lehrach's and G. R. Cantor's laboratories, and on linkage analysis by J. Ott, are well written and informative. But with the sad exception of Huntington's disease, the application of these positional cloning strategies to neurological diseases is almost over. For major psychiatric illnesses, where the mode of inheritance is not clear, strategies other than the lod-score method will probably be used to identify the approximate chromosomal positions of pathogenic loci — the use of positional cloning strategies would seem to be hopelessly naive after the debacles of affective disorder and schizophrenia. Much more likely to succeed are complex genetic analyses of the proportion of genetic risk for disease that is encoded at a candidate locus. This approach is not discussed at all in this book. □

John Hardy is in the Department of Molecular Genetics, St Mary's Hospital Medical School, Norfolk Place, London W2 1PG.

New Journals issue

This year, *Nature's* annual new journals review supplement will appear in the issue of 1 October. Publishers and learned societies are invited to submit journals for review, taking note of the following criteria:

■ Journals that first appeared during or after June 1990 and issued at least four separate numbers by the end of April 1992 will be considered.

■ Journals covering any aspect of science are eligible, although those dealing with clinical medicine, engineering and pure mathematics are excluded, as are publications of abstracts.

■ Frequency of publication must be at least three times a year. The main language used must be English. Translation journals in English are of course eligible.

■ Deadline for submission is the end of May.

When submitting journals for review, please send at least four different issues (the first, the most recent and any two others) of each title.

Catalytic converter

R. Burch

Perspectives in Catalysis. Edited by John M. Thomas and Kirill I. Zamaraev. Blackwell Scientific/IUPAC: 1992. Pp. 492. £69.50.

THIS book is the first in a series to be published by the International Union for Pure and Applied Chemistry under the general heading "Chemistry for the 21st Century". All the chapters offer some new insight into catalysis and some will be of great value. There are a good number of contributions from Eastern Europe, bringing within easy reach some of the excellent work that has been done there. Topics covered are: catalyst design and future developments (three chapters); organometallic and cluster catalysts (three chapters); surface science (one chapter); kinetics and reactor modelling (four chapters); and specific catalysts and catalytic processes (six chapters).

It is invidious, in view of the high overall standard of the contributions, to select particular chapters for attention. But I shall do so to indicate the flavour of the book. The first chapter by J. A. Cusumano on "Creating the future of the chemical industry — catalysts by molecular design" is full of interesting information and should stimulate readers to consider the future of catalytic research. Cusumano first establishes the credentials of catalysis as a technological force and emphasizes the clear economic benefits that are expected to be gained from new catalytic processes. In enlarging upon his theme of development through molecular design, he evokes the picture of catalytic chemists as "masters of molecular design", not quite "masters of the Universe", but pretty close. He argues that over the past decade it has been the advances in six disciplines that have been primarily responsible for moving catalysis into the age of molecular design. These disciplines are theoretical chemistry and simulation, analytical instrumentation, surface science, organometallic chemistry, molecular-sieve science and reaction engineering. The six fields are wide-ranging enough to cover most of catalytic chemistry, so there will be few who would disagree with Cusumano's assessment, although there might be some debate about the relative value of each to catalyst design. It would certainly be appropriate to add materials science and conventional inorganic chemistry as fields that are important to developments in catalyst synthesis.

The chapter by K. I. Zamaraev on

"Perspectives in catalysis: via studies on molecular level to new industrial catalysts and processes" continues the theme of developing new ideas from basic research. For example, surface science studies have provided evidence for surface reconstruction under reaction conditions. Zamaraev emphasizes this point by remarking that if the catalyst is powerful enough to perturb the reaction mixture, the mixture must be equally powerful to perturb the state of the catalyst surface. G. A. Samorjai also discusses adsorbate-induced surface restructuring and the importance of defect sites for bond-breaking reactions. Although surface science has provided evidence in support of this concept, it is important to recall that similar ideas had already been developed many years ago, based on conventional catalytic experiments.

Several chapters include discussions of designer catalysts. A. E. Shilov describes catalysis with organized molecular sys-

tems created to mimic enzyme catalysts; V. A. Likholobov describes the design of catalysts based on metal complexes; and I. I. Moiseev and M. N. Vargaftik discuss catalysis with giant palladium clusters. Zeolite catalysts, as would be expected, feature strongly in the book, with C. S. John, D. M. Clark and I. E. Maxwell providing an excellent overview of new applications of zeolites. Kinetics and reaction modelling are also well covered. In addition, there are chapters on oxides, heteropoly acids, pillared clays, and catalysts for methane conversion, for energy production and for pollution control.

This is a stimulating book that I shall use frequently. Many of the chapters are worth reading more than once, such is the amount of new and interesting information that they contain. □

R. Burch is in the Department of Chemistry, University of Reading, Reading RG6 2AD, UK.

From egg to embryo

Ursula Mittwoch

A History of Regeneration Research: Milestones in the Evolution of a Science. Edited by Charles E. Dinsmore. Cambridge University Press: 1992. Pp. 228. £30, \$54.95.

A Conceptual History of Modern Embryology. Volume 7 of Developmental Biology: A Comprehensive Synthesis. Edited by Scott F. Gilbert. Plenum: 1991. Pp. 266. \$83.40.

How did it come to be accepted that genes transform a fertilized egg into an embryo containing a multitude of cell types and organs? These two books provide a colourful account of research over two-and-a-half centuries that will help developmental biologists put present-day concepts and problems into perspective. Both books are extensively referenced, and there is interdigitation between and within the two.

The term 'genetics' was first used by William Bateson in 1905, and the word 'gene' made its first appearance in a publication by Wilhelm Johannsen in 1909. Genetics was a child of embryology, the study of which evolved in the nineteenth century as an offshoot of the regeneration studies that flourished in the previous century. In *A History of Regeneration Research*, Richard Goss defines regeneration as the growth of amputated structures from an anatomically complex stump. This process, also known as 'epimorphic regeneration', differs from tissue regeneration, or wound healing, not only by its much more

limited distribution, but also in the extent to which lost parts are replaced. In epimorphic regeneration, entirely new structures may be formed even in the absence of their counterparts in the stump.

Among the outstanding names of eighteenth-century regeneration research are those of René-Antoine Ferchault de Réaumur (1683–1757), Abraham Trembley (1710–1784) and Lazzaro Spallanzani (1739–1799), who are the subjects of four fascinating chapters in Dinsmore's book. Réaumur studied regeneration in crayfish, which can cast off limbs that are injured or grasped in a life-threatening situation. He noted that regeneration tended to result in a perfect replacement of the missing part. Although in the wild, limbs were cast off at the basal plate closest to the body, if only the tip of the claw was cut off, this alone would regrow. The underlying mechanism is still puzzling investigators today.

Even more surprising were Trembley's findings on Hydra. If a Hydra was cut transversely into three pieces, the head piece would regenerate a foot, the foot piece a head, while the middle piece would grow a new head and a new foot. But if each part of the animal could regenerate the entire animal, where did the organizing principle reside and how was this principle related to normal development?

A Conceptual History of Modern Embryology opens with an outline of the foundation of classical embryology. Christian Pander (1794–1865) described the germ layers in the developing chick embryo, while Karl Ernst von Baer (1792–1865) related them to later structures in the embryo and the extra-embryonic membrane in different verte-

brates. In 1827, von Baer discovered the mammalian egg, thus ending a search that had begun in the seventeenth century. But he regarded the sperm as parasites, perpetuating his view by naming them "spermatozoa". It was left to Albert Kölliker to trace the life history of the spermatozoa in the testis; their function remained a matter for debate for several more decades. Although the cell theory had been put forward by Mattheus Schleiden and Theodor Schwann in the 1830s, it was not until 40 years later that another generation of investigators, armed with improved microscopes, revealed the secrets of cell division and fertilization. As Frederick Churchill emphasizes, the discovery of chromosomes and their behaviour arose as a refinement of traditional embryological studies.

Investigators were now faced with the problem of the underlying principles of embryonic development. Experiments were conceived that would test the reaction of the embryo to interferences with its normal development, so giving rise to experimental embryology and to a new interest in regeneration.

Among the experiments were those of Wilhelm Roux (1850–1924), who found that when one of the two cells of cleaving frog eggs were punctured with a hot needle, only very few embryos survived as far as the gastrula stage, a finding that he thought lent support to the idea of qualitative cell division. But Hans Driesch (1867–1941) discovered that when he separated blastomeres of cleaving sea-urchin eggs by shaking, they developed into half-sized embryos, some of which reached the larval stage. It seemed, after all that each cell retained "totipotency", enabling it to develop into any part of the organism as the occasion demanded.

There were many experiments in which the effect of calcium on embryonic development was studied. If sea-urchin larvae were grown in calcium-free water, the skeleton did not form, resulting in larvae without arms. This sequence of events suggested to Curt Herbst (1866–1946) the idea of induction, that is, the determination of development of an embryonic region through the influence of another region.

Hans Spemann (1869–1941) was the most famous proponent of induction, and his most celebrated experiment, carried out with Hilde Mangold, was the transplantation of the dorsal lip of the blastopore of a newt larva into the ventral region of another larva, resulting in the development of neural tissue associated with the graft. The results were published in 1924, and in the same year Spemann published another article on "Organizers in animal development".

In 1935, Spemann received a Nobel

prize for elaborating the concept of the organizer, but by then it was already running into difficulties. Johannes Holtfreter, who contributes an autobiographical essay to *A Conceptual History of Modern Embryology*, showed that blastopore lips that had been killed still retained their inducing capacity, which, moreover, seemed to be present in virtually every tissue of every species examined. Soon afterwards, Waddington and Needham found that ether extracts of adult newts had the organizing capacity, and for a time it seemed that the

IMAGE UNAVAILABLE FOR COPYRIGHT REASONS

Thomas Hunt Morgan: made a clear distinction between embryology and genetics.

organizer might be a sterol. But when the dye methylene blue was found to be capable of neural induction, it became evident that the aim of characterizing the organizer was unattainable.

Would the up-and-coming science of genetics be able to explain the mystery of the developing embryo? The contributions of Thomas Hunt Morgan (1866–1945) are discussed in both volumes. Morgan began his scientific career as an embryologist, publishing *The Development of the Frog's Egg: An Introduction to Experimental Embryology and Regeneration* in 1901. Influenced by his friend E. B. Wilson, who with Nettie Stevens was one of the discoverers of sex chromosomes in some species of insects, Morgan studied sex determination in parthenogenetic aphids. In 1908, he began work on the fruitfly *Drosophila melanogaster*, and two years later obtained the first mutant, a white-eyed male who transmitted this characteristic to future generations in a manner that became known as "sex linked", or X-linked, inheritance. In 1913, Alfred Sturtevant, in Morgan's laboratory, published a map of six linked genes on the X chromosome.

Although Morgan retained an active interest in embryology, he made a clear distinction between the subject and genetics, stating in *The Theory of the Gene* (1926) that, whereas genetics deals with the transmission of genes, embryology studies the development of potentialities into adult structures. Since then, genetics has gone from strength to strength, while embryology has apparently languished; but more recently the gap between the two disciplines has narrowed.

In Gilbert's book we learn about three geneticists who set out specifically to bridge the divide. Salome Gluecksohn-Schoenheimer (born in 1907 and still active in research) began work with Spemann in 1928 and emigrated to the United States in 1933, where she obtained a post in L. C. Dunn's laboratory, studying the *T* (brachyury) locus in the mouse. Although her ultimate goal, the elucidation of gene action, eluded her, she played a pioneering role in combining concepts from embryology and genetics to establish developmental genetics. Conrad Hal Waddington's (1905–1975) vision encompassed a very wide field, in which systems, processes and interactions replaced the idea of an inducer acting on a target. He was one of the first to recognize the possible relevance of Jacob and Monod's operator system as a means of generating differential gene action in embryos, linking the induction of bacterial enzymes with the induction of the neural tube. Last, Boris Ephrussi (1901–1979) worked on a large variety of organisms and *in vitro* systems, analysing the role of nuclear and cytoplasmic factors in cell differentiation, thus helping to bring about the transition from classical to molecular genetics, and from embryology to developmental biology.

There is no mention of Hans Grüneberg (1907–1979), who studied the role of genes on skeletal development in mice. One of his findings was that individuals of an inbred strain differed with respect to presence and absence of the third molar tooth, illustrating that the same genotype can give rise to more than one phenotype. As is hinted in Gilbert's book, even the genes may not tell all. □

Ursula Mittwoch is in the Department of Anatomy, Faculty of Basic Medical Sciences, Queen Mary and Westfield College, Mile End Road, London E1 4NS, UK.

New in paperback

■ *Beyond Numeracy: An Uncommon Dictionary of Mathematics* by John Allen Paulos. Published by Penguin, price £6.99.

■ *The Primary Source: Tropical Forests and Our Future* by Norman Myers, now updated for the 1990s. Published by Norton, price \$10.95.

Latest Proterozoic stratigraphy and Earth history

Andrew H. Knoll & Malcolm R. Walter

The end of the Proterozoic Eon was a time of pronounced biological, biogeochemical, climatic and tectonic change. New bio- and chemostratigraphic data provide an improved framework for stratigraphic correlation, making possible a deeper understanding of latest Proterozoic Earth history and providing tools for a chronostratigraphic division of late Proterozoic time.

LIKE the Bible, Earth history is conventionally divided into two segments of markedly different length and character. Earth's 'Old Testament' comprises the Archean and Proterozoic eons: the immense span of Precambrian time that extends from the earliest known rocks to about 540 million years (Myr) ago. It is separated from geology's 'New Testament' (the Phanerozoic Eon) by a discrete event: a unique radiation of eucoelomate animals that permanently changed the nature of ecological and sedimentary systems.

The diverse skeletons that mark the base of the Cambrian System have long presented palaeontologists with an evolutionary problem of the first order. Darwin was keenly aware of it and urged that the sudden appearance of animals in the fossil record was a valid argument against his view of evolution by natural selection¹. He speculated that the Precambrian Earth 'swarmed with living creatures', having in mind principally the animal ancestors of trilobites and other Early Cambrian fauna. Darwin would no doubt be pleased by evidence that microbial life goes back 3,500 Myr or more². But he would be unsettled by the lack of a long Proterozoic record of animal evolution; unequivocal metazoans occur only in the final 40 Myr of the eon, and most of the eucoelomate phyla that populate Phanerozoic seas have no Proterozoic record. The end of the Proterozoic Eon was truly a period of remarkable evolutionary innovation. Less widely appreciated, these biological events are rooted in a broader framework of tectonic, biogeochemical, and climatic change that altered the Earth's surface and, in doing so, ushered in the Phanerozoic world³.

The idea that evolution and environmental change are intertwined is not new, but empirical support for Proterozoic interactions has been difficult to muster because it requires the reliable correlation of physical and biological events documented in the rock record. Students of Phanerozoic history can debate whether extinctions at the Cretaceous-Tertiary boundary took place in a geological instant, precipitated by the impact of a giant bolide. They can link Tertiary vegetational change in North America to climatic deterioration induced by the opening of the Southern Ocean, or ask about the biological consequences of oxygen depletion in late Devonian ocean basins. By contrast, it is not always easy to establish the synchronicity of Archean or Proterozoic events to within 100 million years.

The far better resolution of Phanerozoic history does not stem from a denser framework of radiometric dates. As in the Proterozoic, geochronometry provides a coarse means of correlating Phanerozoic rocks⁴. Fossils and other chronostratigraphical indicators make the difference (see Box 1). The critical importance of chronostratigraphy is most clear in attempts to understand the great events of the late Proterozoic Eon. Important advances have been made in unravelling regional geological histories using geochronometric data⁵, but many sedimentary successions are not easily dated, and even well dated sections cannot be correlated with anything like biostratigraphic precision. In the absence of reliable tools for interbasinal correlation, only a broad understanding of late Proterozoic history has

been attainable. Fortunately, the palaeontological and geochemical records that document this history provide the raw materials for improved stratigraphic correlation. We argue here that, in conjunction with improved geochronometric data, protistan microfossils and isotope geochemistry are likely to provide the means by which the great physical and biological events of the latest Proterozoic Eon can be correlated and, eventually, integrated.

Tradition in Neoproterozoic stratigraphy

Thick sedimentary successions occur immediately beneath Lower Cambrian rocks in many places. Originally referred to as 'Sinian' after exposures along the Yangtze River⁶, these rocks have conventionally been correlated on the basis of two features: glaciogenic sediments and Ediacaran animal remains^{4,7}. Both define broad stratigraphic bands, but each is attended by significant uncertainties.

Palaeozoology. Ediacaran animals comprise a morphologically distinctive fauna of architecturally simple, unskeletalized invertebrates known from over 24 localities on six continents⁸ (Fig. 1b). There is general agreement that Ediacaran assemblages occur only within a discrete interval of latest Proterozoic time, but it is unclear whether all assemblages are strictly coeval, whether all or some taxa extend up to the Proterozoic-Cambrian

BOX 1 Of time and stratigraphy

Geochronology is the discipline in which the ages of rocks are estimated. Estimates can be made in two different ways, and time divisions are, correspondingly, of two sorts. In the chronometric scale, the basic unit is the year and radiometric dating provides the means by which rocks are related to the scale. Chronostratigraphic scales are based on a series of reference points in rock successions, and stratigraphy permits the estimation of age. In the global chronostratigraphic scale, the beginning of each division is defined by a global stratotype section and point (a point in a stratigraphic section chosen by convention to define the boundary between the division in question and the preceding interval (popularly known as the 'golden spike')). Chronostratigraphic age can be known with certainty only in the stratotype section; the ages of all other strata are interpretations based on correlation, the use of the physical, chemical and/or biological features of rocks to establish their age relationships. Correlation may be based on overall characteristics (lithostratigraphy), the specific geometric relationships of sedimentary rock packages (sequence stratigraphy), fossils (biostratigraphy), chemical features (chemostratigraphy) and/or magnetic features (magnetostratigraphy)⁶⁵. Ideally, correlation relies on the recognition in different successions of geological signatures judged to be synchronous (of the same age wherever detected). In fact, many signatures, including most biostratigraphic and climatological events, are at least slightly diachronous; that is, they differ in age from place to place. The degree to which diachroneity compromises correlation depends on the stratigraphic resolution required. Like phylogenetic hypotheses in biology, hypotheses of correlation may be well or poorly corroborated. The well known geological timescale for the Phanerozoic Eon is a chronostratigraphic scale that has been integrated with the chronometric scale by means of geochronological calibration.

boundary, or whether distinct biozones can be recognized within the interval. It is likely that assemblage zones will eventually be established, but for the present, such questions can be addressed most effectively using stratigraphic tools other than the fossils themselves. Taphonomic, palaeoenvironmental and, possibly, biogeographical variations further complicate the picture in ways that remain poorly understood⁸. Perhaps most limiting, Ediacaran body fossils are relatively rare.

Trace fossils provide additional and largely independent evidence of Proterozoic animal evolution (Fig. 1c). The assemblages of simple tracks, traces, and burrows found in Proterozoic rocks are distinct from those in basal Cambrian and younger deposits, and they seem to be recognizable globally in siliciclastic sediments⁹. Thus, trace fossils provide useful markers of the Proterozoic–Cambrian boundary and may allow as many as three uppermost Proterozoic biozones to be recognized^{10–12} (Fig. 3).



FIG. 1 Latest Proterozoic metazoans. *a*, *Cloudina hartmanni*, a skeletalized animal from the Nama Group, Namibia, in longitudinal and transverse sections; $\times 5$ (photograph by Stephen Grant). *b*, *Dickinsonia costata*, a non-skeletal, Ediacara-grade metazoan from the Pound Subgroup, Adelaide Geosyncline, South Australia; $\times 1.3$ (photograph by R. J. F. Jenkins). *c*, *Planolites ballandus*, a trace fossil from the Elkeru Formation, Georgina Basin, Australia; $\times 1$.

Conventionally, the evolution of skeletons is seen as the defining event of the Proterozoic–Cambrian boundary, but in Namibia, carbonates interbedded with sandstones bearing Ediacaran fossils contain complex tubular skeletons made of calcite^{13,14} (Fig. 1a). Indeed, architecturally similar skeletons occur throughout the world in carbonates that are definitely or probably of latest Proterozoic age^{14,15}. The degree of stratigraphic overlap between Ediacaran fossils and these early skeletons is uncertain. Chitinous sabelliditid worms and the first biomineralized shells of Cambrian aspect appear at or just below the Proterozoic–Cambrian boundary¹⁶.

Invertebrate animals provide a reliable guide to the broad recognition of uppermost Proterozoic strata, but because we can at present correlate using only a grade of evolution rather than well defined species ranges, invertebrate biostratigraphy remains a fairly coarse means of correlating Neoproterozoic strata. (A recent report by Hofmann *et al.*¹⁷ of possible metazoan impressions in older rocks opens up the possibility that an earlier stratigraphic interval may be defined by simple animals, but it does not alter the general conclusion.)

Neoproterozoic ice ages. Tillites and associated glaciogenic strata occur throughout the world in upper Proterozoic successions¹⁸. In 1964, Harland¹⁹ hypothesized that these record a great late Proterozoic ice age of global extent. It has now become clear that the tillites vary in age between about 850 and 590 Myr¹⁸. Opinion has varied as to the discreteness of Neoproterozoic ice ages; at present, many would agree that there were at least two major glacial periods and probably more. Within each ice age, deposits are broadly but not necessarily strictly coeval. Of particular interest has been the Varanger (=Laplandian; ~ 610 –590 Myr) ice age, probably the most widespread of the Neoproterozoic glaciations. Varanger tillites occur stratigraphically below all known diverse Ediacaran faunal assemblages. As with metazoan fossils, the extent of diachroneity among Neoproterozoic glaciogenic rocks must be evaluated using independent methods of correlation.

Stromatolites. Like invertebrates, microbial communities have left a record of trace fossils, in this case the distinctively laminated sedimentary structures known as stromatolites. Coarse patterns of temporal change have been recognized in the Proterozoic stromatolite record²⁰, and recent observations continue to suggest that these structures can have a useful auxiliary role in stratigraphic correlation^{16,21–23}. But there is no evidence that stromatolites can resolve the stratigraphic uncertainties noted above or match the resolution discussed below²⁴.

Emerging methodologies

Fossil protists and prokaryotes. Photosynthetic organisms are abundantly represented in uppermost Proterozoic rocks, but their stratigraphic potentials vary widely. Seaweeds can be morphologically complex, but their low probability of preservation results in a patchy record of uncertain stratigraphic usefulness. One possible exception may be the problematic organic ribbons known as vendotaenids, which regularly occur in uppermost Proterozoic sediments²⁵. But even these define a broad stratigraphic band whose boundaries remain uncertain. Cyanobacteria and cyanobacteria-like microfossils are widely distributed in uppermost Proterozoic rocks, but virtually all have close similarities both to older fossil and to living taxa. Most distinctive is the helical *Obruchevella*, which has a distinct (but stratigraphically broad) acme in uppermost Proterozoic and Lower Cambrian rocks.

The most promising fossils for Neoproterozoic biostratigraphy are the acritarchs, organic-walled microfossils produced for the most part by phytoplanktonic protists. Acanthomorphic and other morphologically complex acritarchs have been useful in correlating Phanerozoic successions but, until recently, most palaeontologists considered such fossils to be rare or absent from Proterozoic successions. Now, over 24 assemblages of complex acritarchs have been described from Neoproterozoic

rocks, many of these fossils being extremely large relative to younger examples (Fig. 2). Although the total number of assemblages remains small, it is becoming clear that several biostratigraphic zones can be recognized (Fig. 3). Varanger and immediately post-Varanger acritarchs can be distinguished from earlier assemblages²⁶⁻³⁰, and further subdivision may be possible. Most species of large, morphologically complex acritarchs disappeared at about the time of the Ediacaran faunal radiation, leaving very latest Proterozoic assemblages characteristically simple and low in diversity¹⁶. Several distinctive taxa appeared near the very end of the Proterozoic, and a major radiation of acritarchs marks the beginning of the Cambrian^{16,31}. Thus, microfossil assemblages provide an independent framework for evaluating the stratigraphic ranges of early animals and for correlating the many Neoproterozoic successions in which no metazoan fossils are known.

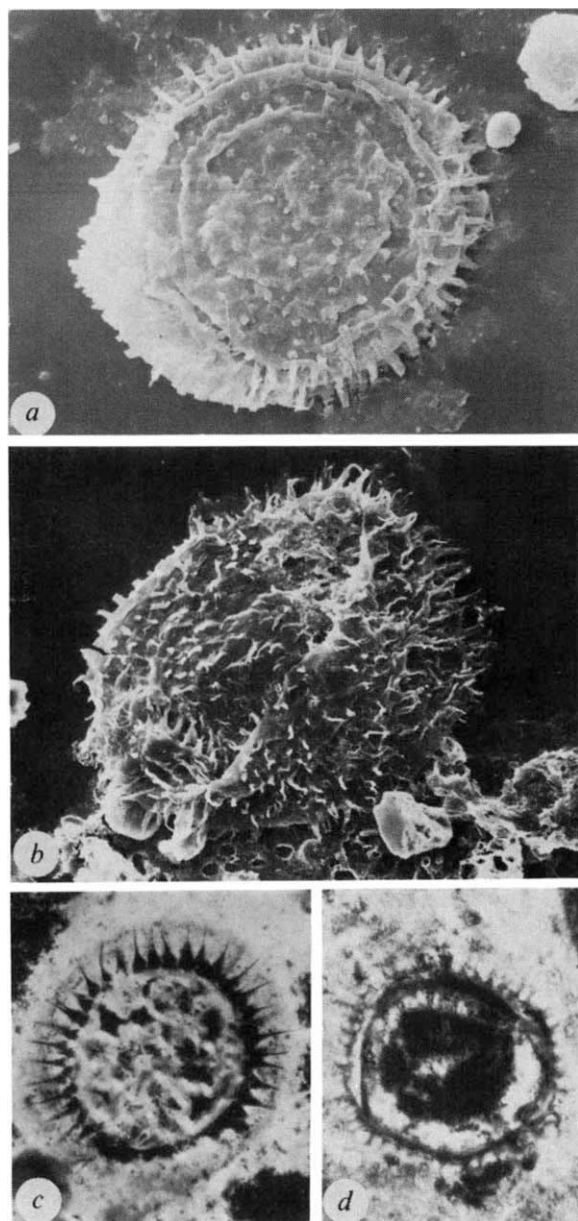


FIG. 2 Morphologically complex acritarchs from terminal Proterozoic successions. *a, b*, From the Pertatataka Formation, Amadeus Basin, Australia; vesicle diameters 155 and 200 μm , respectively (photographs by Zang Wenlong). *c*, From the Doushantou Formation, China; vesicle diameter 200 μm (material from Zhang Yun). *d*, From greenschist-facies metasediments of the Scotia Formation, Prins Karls Froland, Svalbard; vesicle 155 μm .

Isotopic chemostratigraphy. Secular variations in the isotopic composition of sea water are useful in the correlation of Phanerozoic successions. For example, oxygen isotopic variation linked to the growth and decay of ice sheets has been critically important in unravelling Pleistocene history. A pronounced and essentially monotonic increase in $^{87}\text{Sr}/^{86}\text{Sr}$ over the past 100 Myr provides stratigraphic resolution that rivals biostratigraphy. More broadly, the C and Sr isotopic compositions of carbonates and the S isotopic composition of sulphates exhibit stratigraphically useful patterns of secular variation throughout the Phanerozoic³².

A common prejudice is that although Phanerozoic carbonates may be little disturbed, their Proterozoic counterparts are almost invariably altered. There is little empirical evidence to support this supposition. On the contrary, a number of studies have demonstrated that depositional C and Sr isotopic signatures can be retained in micritic carbonates and oolites bound by fine-grained syndimentary cements³³⁻⁴¹. A principal finding is that these signatures varied markedly during Neoproterozoic time, providing a secular pattern of strong stratigraphic potential (Fig. 3). Detailed isotopic curves for the Neoproterozoic Eon are being constructed from analyses of numerous rock samples collected at close intervals in individual measured sections, with sedimentological and petrographic control.

The carbon isotopic composition of carbonates and organic matter formed in sea water reflects both the degree of isotopic fractionation by photoautotrophs and carbon fluxes into and out of the ocean. Rates of organic carbon burial exert a primary influence on these ratios, with higher $\delta^{13}\text{C}$ indicating greater rates of burial; ocean mixing can also be significant. At the close of the Varanger ice age, both carbonate and organic carbon isotopic ratios plunged to distinctly light values ($\delta^{13}\text{C}_{\text{CARB}} = -2$ to -4% relative to the PDB Standard)³³⁻³⁵. A subsequent positive excursion resulted in a short-lived peak of about $+4$ to $+5\%$, after which $\delta^{13}\text{C}_{\text{CARB}}$ remained at about $+2\%$ in latest Proterozoic time^{35,36}. Profiles from Siberia and elsewhere indicate another rapid shift to negative and then strongly positive values just below the Proterozoic-Cambrian boundary, with basal Cambrian $\delta^{13}\text{C}_{\text{CARB}}$ running to about -1% (ref. 42).

Fewer pre-Varanger profiles have been examined, but successions in Svalbard, East Greenland, northwestern Canada and Namibia show similarly distinctive patterns of secular variation^{33,35,40}. Many 850-610 Myr carbonates contain carbon that is isotopically heavy ($\delta^{13}\text{C}_{\text{CARB}} \geq 5\%$) and matched by correspondingly heavy organic carbon; this signal is punctuated by several shorter intervals of isotopically light carbonate (about 0 to -2%). It might be argued that these negative excursions are of diagenetic origin; for example, carbonates exposed sub-aerially during low sea level stands commonly become altered by isotopically light ground waters⁴³. But known isotopic excursions are associated with transgression rather than regression, and organic carbon signatures co-vary smoothly with the carbonates, an unlikely circumstance if the signals were produced diagenetically. Several and perhaps all of the negative excursions fall at the top of glaciogenic horizons³⁵. It seems, therefore, that the marked isotopic variations of the interval 850-610 Myr reflect changes in the world ocean and as such provide a useful stratigraphic signal.

The Neoproterozoic carbon isotopic record thus forms an oscillating pattern that, like the Pleistocene oxygen isotope curve, must be paired with other stratigraphic tools to achieve its full potential. Accumulating evidence also makes it clear that despite the common preservation of near primary carbon isotopic ratios in Neoproterozoic carbonates and organic matter, every sample must be screened for possible diagenetic alteration³⁷.

A second isotopic system of stratigraphic value is that of strontium. The Sr isotopic composition of sea water (and of carbonates precipitated from sea water) reflects the relative contributions of continental erosion (generally high $^{87}\text{Sr}/^{86}\text{Sr}$,

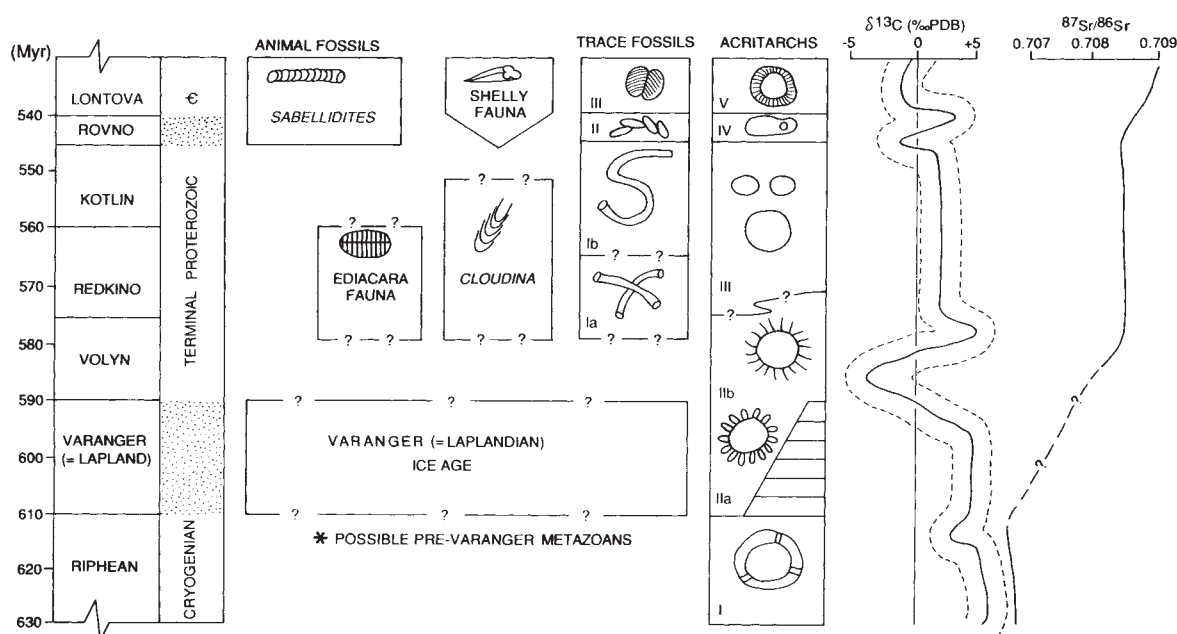


FIG. 3 Summary of terminal Proterozoic biostratigraphy and chemostratigraphy, showing the position of the Varanger ice age. Stratigraphic names are modified from Sokolov⁶²; the shaded areas between periods indicate that global stratotype sections and boundary points have yet to be ratified by the International Union of Geological Sciences. Geochronometric estimates

are best guesses, and are subject to change as data improve. The numbered vertical sequences of trace fossils and acritarchs indicate biostratigraphic zones that can be used for the intercontinental correlation of sedimentary successions. C and Sr isotopic curves are based on the analysis of carbonate rocks. See text for discussion and data sources.

but varying as a function of source rock) and hydrothermal input (typically low $^{87}\text{Sr}/^{86}\text{Sr}$). Like C isotopes, the Sr isotopic compositions of Neoproterozoic carbonates define a curve of great geochemical interest and stratigraphic usefulness (Fig. 3). As first argued by Veizer *et al.*⁴⁴, $^{87}\text{Sr}/^{86}\text{Sr}$ is remarkably low in ~850–800 Myr carbonates; values as low as 0.7055 reflect a relative hydrothermal contribution greater than anything during the succeeding 800 Myr⁴⁰. Ratios then increased to ~0.7070, where they remained until the Varanger glaciation³⁹. By the end of the Varanger ice age, $^{87}\text{Sr}/^{86}\text{Sr}$ had increased to 0.7080, and by the time that Ediacaran metazoans diversified, the ratio stood at 0.7085 (refs 36, 44). Strontium isotopic ratios remained at about that level until near the Proterozoic–Cambrian boundary, when another increase lifted $^{87}\text{Sr}/^{86}\text{Sr}$ to values approached only by the present^{36,45,46}. Thus, secular change in the Sr isotopic composition of sea water provides another chemostratigraphic marker, and one whose stratigraphic sensitivity is potentially fine for certain periods, particularly the intervals near the Varanger glaciation and the Proterozoic–Cambrian boundary.

The isotopic composition of sulphur in Neoproterozoic sulphates is also of potential stratigraphic interest insofar as a major shift in $\delta^{34}\text{S}$ occurred near the end of the Proterozoic⁴⁷. Unfortunately, the stratigraphic sampling of Neoproterozoic sulphates is patchy so that this pattern remains rather broadly defined. Sulphide S is ubiquitous in Neoproterozoic sedimentary rocks, but its isotopic composition reflects local sulphate availability and bears no simple relationship to sea water composition. This problem may potentially be overcome by limiting stratigraphic interpretation to those sulphides that reflect maximum isotopic fractionation (see for example ref. 48), but few systematic data are available.

Sequence stratigraphy. Stratigraphers have long used transgressive and regressive facies in lithostratigraphic correlation. In recent years, a more rigorously articulated analysis of sedimentary sequences has provided potentially fine chronostratigraphic information for the correlation of Phanerozoic successions; sequence boundaries seem to be synchronous on a regional scale, at the resolution of other stratigraphic tech-

niques⁴⁹. Although it has been proposed that eustatic change of sea level exerts a primary influence on sequence development, the controls are more complex and include local tectonism.

Packages of genetically related sedimentary facies bounded by unconformities or their correlative conformable surfaces can be recognized, as well, in Proterozoic successions^{50–55}. Because sequence boundaries share sets of common characteristics that recur through time, independent criteria must be used to enable (or at least support) the inter-regional correlation of sequences. There has been far too little research to allow the confident recognition of Proterozoic sequences likely to reflect eustatic sea level change; however, there are intriguing hints. Glaciations inevitably lead to changes of sea level, and ice-age-related sequences that may well be of global significance have been recognized. For example, the sequence overlying the upper (Marinoan, probably equivalent to Varanger) glacial beds in the Adelaide Geosyncline, Australia, is lithologically distinctive⁵² and has long been correlated across the continent (for example ref. 56) on the basis of stratigraphic position and a distinctive marker bed (the 'upper marker cap dolomite'). Marker beds that are the product of unusual biological and/or environmental conditions may enable otherwise similar sequences to be distinguished.

A special complication arises in Proterozoic analyses because many historically important successions formed in intracratonic settings. In such settings, depositional space is available only intermittently, resulting in a very incomplete sedimentary record that may be difficult to correlate between regions.

Although it may be premature to correlate sequence boundaries among continents, sequence analysis within basins is necessary for the evaluation of bio- and chemostratigraphic data, because stratigraphic patterns in basins will be influenced by regional patterns of non-deposition or erosion, as well as by global secular changes.

Geochronometry. The International Commission on Stratigraphy recommends that newly selected stratotypes for system boundaries be amenable to radiometric dating. This is clearly desirable, as is the dating of significant events within systems,

but it is rarely easy to achieve. The history of attempts to date the initial Cambrian boundary illustrates this. For the past 20 years, age estimates have ranged from 530 to 600 Myr, an uncertainty equal to the full duration of the Ordovician Period⁵⁷. The older estimates come from various techniques for the direct dating of sedimentary rocks, including K/Ar on glauconite and illite and Rb/Sr on shale, and to a lesser extent from K/Ar and Rb/Sr determinations on igneous rocks. Dating igneous rocks can provide precise and accurate results, but it can be difficult to establish the relationship of the dated event to sedimentary rocks of interest. The best lithologies for the calibration of Neoproterozoic time are volcanic ash beds, rocks that occur in a clear stratigraphic context and contain minerals such as zircon that can be dated precisely by U/Pb techniques.

Even zircon dating requires care, because zircon is a recalcitrant mineral that can be weathered from its host rock, redeposited and later metamorphosed, all without disturbing its isotopic composition and thus the age will yield. Both sedimentary and igneous rocks commonly contain several generations of zircon grains, and these must be dated separately if a meaningful age is to be obtained. A powerful new approach has been made possible by the development of ion microprobes that can analyse small parts of single zircon grains⁵⁸. Individual zircon populations can be identified, permitting meaningful and precise age determinations. Even so, if only a single zircon population is present in a sample, it can be difficult or impossible to determine whether it should be considered pristine or inherited, and thus what any resulting date might signify. Radiometric dates on igneous rocks may be supplemented usefully by U/Pb and/or Pb/Pb dates on early diagenetic phosphorites and carbonates.

At present, evidence favours a Proterozoic–Cambrian boundary at ~540 Myr^{58–60} and a range for the Varanger ice age of about 610–590 Myr^{4,60}.

A timescale for the Neoproterozoic Era

The Phanerozoic geological timescale both reflects and is characterized by historical events, chiefly biological. By contrast, a recently ratified convention imposes a strictly chronometric basis for subdividing the Archean and Proterozoic eons⁶¹. This refinement of the timescale provides coarse stratigraphic tools, but the study of Earth history demands a well honed scalpel. The significant biological, climatic and chemostratigraphic events of the Neoproterozoic Era suggest that the chronostratigraphic basis of Phanerozoic correlation can be extended downward to cover at least the last 300 Myr of the Proterozoic Eon.

In chronostratigraphic timescales, a division is defined by a point in a global stratotype (see Box 1). Boundary points are chosen by convention, but their placement depends on conceptions of what they are meant to bound and the prospects for correlation from the stratotype section to other parts of the globe. Several chronostratigraphic divisions have been proposed for latest Proterozoic time. Sokolov⁶² proposed a Vendian Period, with an initial boundary originally based on a widespread transgressive sequence in Ukraine that bears an Ediacara fauna, but later broadened to include Varanger tillites. Jenkins⁶³ and Cloud and Glaessner⁶ independently proposed an Ediacaran (or Ediacarian) Period, with initial boundary stratotypes in South Australia placed to reflect the first appearance of Ediacaran metazoans and the post-glacial (Marinoan, probably equivalent to Varanger) transgression, respectively. Harland and Herod⁶⁴ advocated a Vendian Period containing an early Varanger and later Ediacara epoch (terminology as in ref. 4).

We agree that latest Proterozoic Earth history may best be characterized in terms of ice ages and animal fossils, but argue that the stratigraphic tools that will unlock this history are microfossils and stable isotopes. Used together, acritarchs, C isotopes and Sr isotopes permit correlation between Neoproterozoic successions in Svalbard and western Canada at the formation level⁴⁰. Similarly fine correlations are proving

possible among uppermost Proterozoic basins as widely scattered as Australia, Siberia, southern China, Africa and the lesser Himalaya. As palaeontological and isotopic data are collected simultaneously and evaluated in light of the sequence stratigraphy of basins under consideration, Neoproterozoic correlation will come of age.

The potential for improved correlation has implications for the choice of stratotype section and boundary point for a terminal Proterozoic division. As noted above, several proposals place the beginning of this unit at the onset of Varanger tillite deposition. But the beginning of this ice age is difficult to identify on grounds other than the appearance of glaciogenic sediments. Given the uncertainties regarding tillite synchronicity, global correlation from a stratotype section so chosen is likely to prove difficult. By contrast, distinctive microfossil assemblages, a pronounced transgression, a major decrease in $\delta^{13}\text{C}$, and the rise of $^{87}\text{Sr}/^{86}\text{Sr}$ to 0.7080 all fall at or near to the end of the Varanger ice age. We therefore suggest that the initial boundary of a terminal Proterozoic division be defined by a boundary stratotype that reflects these attributes. This has the advantage of a chronostratigraphic definition based on events that can be identified globally in a wide range of lithologies. The best stratotype section is yet to be determined, but candidates include the Adelaide Geosyncline, Australia; the Windermere Supergroup, northwestern Canada; the Duoshantuo/Denying succession, China; the Krol succession in the lesser Himalaya; the Witlvei and Nama groups, Namibia; and several successions in Siberia. By definition, the end of this interval coincides with the beginning of the Cambrian Period.

A second advantage of such a definition is that it would make the Varanger glaciation the final event of a previous division. Ice ages, iron formation, distinctive protistan fossils and stromatolites, unusual C isotope profiles, and unique Sr isotopic ratios collectively mark the time interval from ~850 to 590 Myr as one of the most distinctive in the past 2,000 Myr. The International Commission on Stratigraphy⁶⁰ has recognized a Cryogenian Period defined chronometrically as beginning 850 Myr ago. This definition needs to be complemented by a chronostratigraphic scale based on microfossils and isotopic geochemistry, with a major effort made to tie such data (as well as those of the terminal Proterozoic interval) into the broader chronometric framework of the Proterozoic timescale.

Conclusion

We stand at the edge of a new understanding of late Proterozoic Earth history. Previous research has demonstrated that this was a remarkable interval characterized by profound biological, tectonic and environmental events. The challenge now is to use new methods of stratigraphic correlation to integrate these disparate records so that we may learn how the foundations for Phanerozoic physical and biological evolution were established. The emerging stratigraphic framework promises a new exegesis of the eventful final chapters in Earth's Old Testament. □

Andrew H. Knoll is at the Botanical Museum, Harvard University, Cambridge, Massachusetts 02138, USA; Malcolm R. Walter is at the School of Earth Sciences, Macquarie University, New South Wales 2109, Australia.

1. Darwin, C. R. *The Origin of Species* (Murray, London, 1859).
2. Schopf, J. W. (ed.) *Earth's Earliest Biosphere: Its Origin and Evolution* (Princeton Univ. Press, Princeton, 1983).
3. Knoll, A. H. *Scient. Am.* **265**, 65–73 (1991).
4. Harland, W. B. et al. *A Geological Time Scale 1989* (Cambridge Univ. Press, Cambridge, 1990).
5. Krogh, T. E., Strong, D. F., O'Brien, S. J. & Papezik, V. S. *Can. J. Earth Sci.* **25**, 442–453 (1988).
6. Willis, B., Blackwelder, E. & Sargent, R. H. *Research in China* (Carnegie Institute of Washington, 1907).
7. Cloud, P. & Glaessner, M. F. *Science* **217**, 783–792 (1982).
8. Glaessner, M. F. *The Dawn of Animal Life* (Cambridge Univ. Press, Cambridge, 1984).
9. Crimes, T. P. *Geol. Mag.* **124**, 97–119 (1987).
10. Fedorkin, M. A. in *The Vendian System: Vol. 1. Paleontology* English edn (ed Solovov, B. S. & Iwanowski, M. A.) 112–117 (Springer, Berlin, 1990).
11. Narbonne, G. M. & Myrow, P. *N.Y. St. Mus. Bull.* **463**, 72–76 (1988).
12. Walter, M. R., Elphinstone, R. & Heys, G. R. *Alcheringa* **13**, 209–256 (1989).
13. Germs, G. J. B. *Am. J. Sci.* **272**, 752–761 (1972).

14. Grant, S. *Am. J. Sci.* **290**, 261–294 (1990).
15. Conway Morris, S., Mattes, B. W. & Chen, M. *Am. J. Sci.* **290A**, 245–260 (1990).
16. Sokolov, B. S. & Iwanowski, A. B. (eds) *The Vendian System: Vol. 1 Paleontology* English edn (Springer, Berlin, 1990).
17. Hofmann, H., Narbonne, G. M. & Aitken, J. D. *Geology* **18**, 1199–1202 (1990).
18. Hambrey, M. B. & Harland, W. B. (eds) *Earth's Pre-Pleistocene Glacial Record* (Cambridge Univ. Press, Cambridge, 1981).
19. Harland, W. B. *Geol. Rundschau* **54**, 45–61 (1964).
20. Walter, M. R. (ed.) *Stromatolites* (Elsevier, Amsterdam, 1976).
21. Bertrand-Sarfati, J. & Walter, M. R. *Precambrian Res.* **15**, 353–371 (1981).
22. Walter, M. R. & Heys, G. R. *Precambrian Res.* **29**, 149–175 (1985).
23. Hofmann, H. J. *Geosci. Canada* **14**, 135–154 (1987).
24. Khomentovsky, A. A. in *The Vendian System Vol. 2 Regional Geology* English edn (eds Sokolov, B. S. & Fedonkin, M. A.) 102–183 (Springer, Berlin, 1990).
25. Gnilovskaya, M. B. (ed.) *Vendotaenids of the East European Platform* (Nauka, Leningrad, 1988).
26. Yin, L. *Palaeontologia Cathayana* **2**, 229–249 (1985).
27. Zang, W. & Walter, M. R. *Nature* **337**, 632–645 (1989).
28. Knoll, A. H. & Butterfield, N. J. *Nature* **337**, 602–603 (1989).
29. Vidal, G. *Palaeontology* **33**, 287–298 (1990).
30. Knoll, A. H. *Palaeontology* (in the press).
31. Moczyłowska, M. *Fossils Strata* **29**, 1–127 (1991).
32. Holser, W. T. in *Patterns of Change in Earth Evolution* (eds Holland, H. D. & Trendall, A. F.) 123–143 (Springer, Berlin, 1984).
33. Knoll, A. H., Hayes, J. M., Kaufman, A. J., Swett, K. & Lambert, I. B. *Nature* **321**, 832–838 (1986).
34. Fairchild, I. R. & Spiro, B. *Sedimentology* **34**, 973–988 (1987).
35. Kaufman, A. J., Hayes, J. M., Knoll, A. H. & Germs, G. J. B. *Precambrian Res.* **49**, 301–327 (1991).
36. Lambert, I. B., Walter, M. R., Zang, W., Lu, S. & Ma, G. *Nature* **325**, 140–142 (1987).
37. Fairchild, I. R., Marshall, J. D. & Bertrand-Sarfati, J. *Am. J. Sci.* **290A**, 46–79 (1990).
38. Aharon, P., Schidlowski, M. & Singh, I. B. *Nature* **327**, 699–702 (1987).
39. Derry, L. A., Keto, L. S., Jacobsen, S. B., Knoll, A. H. & Swett, K. *Geochim. cosmochim. Acta* **53**, 2331–2339 (1989).
40. Asmerom, Y., Jacobsen, S. B., Knoll, A. H., Butterfield, N. J. & Swett, K. *Geochim. cosmochim. Acta* **55**, 2883–2894 (1991).
41. Derry, L. A., Kaufman, A. J. & Jacobsen, S. B. *Geochim. cosmochim. Acta* (in the press).
42. Kirschvink, J. L., Magaritz, M., Ripperdam, R. L., Zhuravlev, A. Y. & Rozanov, A. Y. *GSA Today* **1**, 69–91 (1991).
43. Beanus, M. A. & Knauth, L. P. *Geol. Soc. Am. Bull.* **96**, 737–745 (1985).
44. Veizer, J., Compston, W., Clauer, N. & Schidlowski, M. *Geochim. cosmochim. Acta* **47**, 295–302 (1983).
45. Donnelly, T. H., Shergold, J. H., Southgate, P. N. & Barnes, C. N. in *Phosphorite Research and Development* (ed. Notholt, A. J. G. & Jarvis, I.) *Geol. Soc. Spec. Publ.* **52**, 273–287 (1990).
46. Kaufman, A. J., Jacobsen, S. B. & Knoll, A. H. *Earth planet. Sci. Lett.* (in the press).
47. Claypool, G. E., Holser, W. T., Kaplan, I. R., Sakai, H. & Zak, I. *Chem. Geol.* **28**, 199–260 (1980).
48. Ross, G. M., Bloch, J. D. & Krouse, H. R. *Abstr. Geol. Assoc. Canada/Miner. Assoc. Canada Ann. Mtg* **16**, 108 (1991).
49. Christie-Blick, N. *Mar. Geol.* **97**, 35–56 (1991).
50. Lindsay, J. F. *Am. Assoc. Petrol. Geol. Bull.* **71**, 1387–1403 (1987).
51. Lindsay, J. F. & Korsch, R. J. *Basin Res.* **2**, 3–25 (1989).
52. von der Borch, C. C., Christie-Blick, N. & Grady, A. E. *Austral. J. Earth Sci.* **35**, 59–72 (1988).
53. Christie-Blick, N., Grotzinger, J. P. & von der Borch, C. C. *Geology* **16**, 100–104 (1988).
54. Christie-Blick, N., von der Borch, C. C. & DiBona, P. A. *Am. J. Sci.* **290A**, 295–332 (1990).
55. Ross, G. M. & Murphy, D. C. *Geology* **16**, 139–143 (1988).
56. Preiss, W. V., Walter, M. R., Coats, R. P. & Wells, A. T. *BMJ J. Aust. Geol. Geophys.* **3**, 43–53 (1978).
57. Cowie, J. W. & Harland, W. B. in *The Precambrian–Cambrian Boundary* (eds Cowie, J. R. & Brasier, M. D.) 186–204 (Clarendon, Oxford, 1989).
58. Compston, W. et al. *Res. Sch. Earth Sci. Aust. Nat. Univ. A. Rep.* **1990**, 26–27 (1990).
59. Jenkins, R. J. F. *Geol. Mag.* **121**, 635–643 (1984).
60. Conway Morris, S. in *The Precambrian–Cambrian Boundary* (eds Cowie, J. R. & Brasier, M. D.) 7–39 (Clarendon, Oxford, 1989).
61. Plumb, K. *Episodes* **14**, 139–140 (1991).
62. Sokolov, B. S. *Izv. Akad. nauk SSSR ser. geol.* **5**, 21–31 (1952).
63. Jenkins, R. J. F. *Trans. R. Soc. S. Aust.* **105**, 179–194 (1984).
64. Harland, W. B. & Herod, K. M. in *Ice Ages: Ancient and Modern* (eds Wright, A. E. & Moseley, F.) 189–216 (Steel Horse, Liverpool, 1975).
65. Whittaker, A. et al. *J. geol. Soc. Lond.* **148**, 813–824 (1991).

ACKNOWLEDGEMENTS. We thank L. Smith, S. Grant, S. Conway Morris, P. Link, W. B. Harland, J. Veevers and R. Jenkins for comments. This work was supported by NASA, NSF and a Visiting Fellowship from Gonville and Caius College, Cambridge (A.H.K.) and the Australian Research Council (M.R.W.)

ARTICLES

Upper-mantle seismic discontinuities and the thermal structure of subduction zones

John E. Vidale & Harley M. Benz

United States Geological Survey, Branch of Seismology, 345 Middlefield Road, MS 977, Menlo Park, California 94025, USA

The precise depths at which seismic velocities change abruptly in the upper mantle are revealed by the analysis of data from hundreds of seismometers across the western United States. The boundary near 410 km depth is locally elevated, that near 660 km depressed. The depths of these boundaries, which mark phase transitions, provide an *in situ* thermometer in subduction zones: the observed temperature contrasts require at least moderate thickening of the subducting slab near 660 km depth. In addition, a reflector near 210 km depth may mark the bottom of the aesthenosphere.

OUR understanding of the composition, convective pattern and temperature field in the Earth's mantle is derived from many kinds of geophysical evidence. The radial seismic structure is an important clue to the dynamics of the mantle. At least two depths in the mantle are known from seismology to be the location of transitions in rock properties. Near 410 km depth, the transition from the α to the β phase of olivine (hereafter referred to as the '410') is identified by a 5% increase in shear and compressional velocity and density. Near 660 km depth (hereafter the '660'), the P and S velocities and density again increase by ~5%, consistent with the transition from γ -phase olivine to perovskite plus magnesiowüstite^{1,2}, and perhaps an increase in iron and silica content.

There is continued debate as to whether the mantle convects in a single system from crust to core, or whether mantle convection is stratified, with the '660' inhibiting mass transport between the upper and lower mantle^{2,3}. Many details remain unresolved. Measurement of the sharpness, contrast and depth variations of both the '410' and '660' boundaries can contribute to estimating the lateral temperature contrasts as well as the composition of the mantle.

Measurements of the '410' and '660' boundaries have been made with a variety of seismic techniques. These boundaries generate reflections and conversions, as well as more complicated arrivals near the critical angle. Generally, long-period studies measure boundary depths and impedance contrasts more precisely, whereas short-period studies have greater resolution of short-wavelength features such as sharpness and lateral depth variation. The '410' and '660' were discovered from studies of seismic body waves recorded at distances between 10 and 35° from earthquakes. Within part of this range, for each boundary, it is possible to observe an arrival that travelled above the discontinuity, an arrival that was reflected by the discontinuity and an arrival that penetrated below the discontinuity⁴.

Global observations of the '410' and '660' discontinuities have been obtained from reflected waves of 25–50 s period^{2,5,6}. The lateral resolution of these waves is ~1,000 km. The temperature field near subducting slabs, however, is expected to show shorter-wavelength variations, because of the ~100-km scale length of thermal diffusion of lithosphere subducted 10 million years ago. Temperature variations are expected to result in variations in the depth of both the '410' and '660' boundaries. Consequently,

shorter-period seismic waves are necessary to measure these shorter-wavelength boundary deflections.

In contrast to previous work, here we use a large-aperture array with many elements, each of which is a short-period seismometer. Array study of the waveforms improves the signal-to-noise ratio by the cancellation of the near-receiver reverberations and other noise, which differ for each station. We examine one earthquake at a time to find mantle reflections that leave the source region in a fairly narrow cone of take-off angles. Our measurements sample mantle structures across a lateral extent of only a few tens of kilometres, giving much higher resolution than previous work.

The large-aperture dense array

The five regional, short-period arrays whose seismograms we examine are the Northern California Seismic Network, the Southern California Seismic Network, the University of Utah Regional Seismic Network, the University of Nevada-Reno Seismic Network and the Washington Regional Seismic Network. Combined, the arrays consist of 881 short-period vertical stations across the western United States (Fig. 1). The resulting antenna, which spans 1,700 km by 800 km, covers more than 30 times the area of the short-period arrays previously used for mantle waveform studies. From six earthquakes, we observe mantle reflectors beneath South America, Tonga, Bonin and Mariana with roughly the geometry shown in Fig. 2. These earthquakes are recorded with good signal-to-noise ratios at frequencies between 0.2 and 5.0 Hz.

With the spacious aperture and the multiplicity of stations, we can effectively eliminate the largest sources of noise, which are microtremor, electronic noise and near-station reverberations. The scattering near the source is minimized by considering the part of the wavefield that arrives before pP, the earliest arriving energy scattered off the Earth's surface or ocean bottom above the source.

The seismograms that we analyse contain most of their energy at a period of 4 s. Although the instruments are designed to measure shorter-period seismic energy, they have been used to study ground motion up to 20–30 s period⁷. We visually examine every trace and reject the 40–70% that are too noisy or dissimilar to nearby stations. A time in the P or pP arrival is handpicked, either the onset time or a prominent early peak, depending on

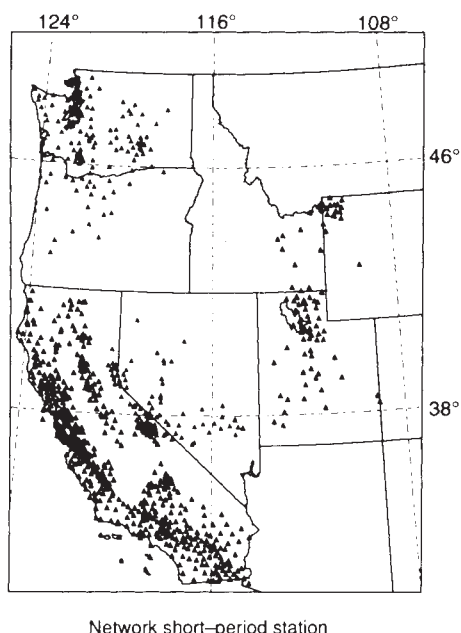


FIG. 1 Location of the 881 stations in the five regional short-period arrays assembled in this study.

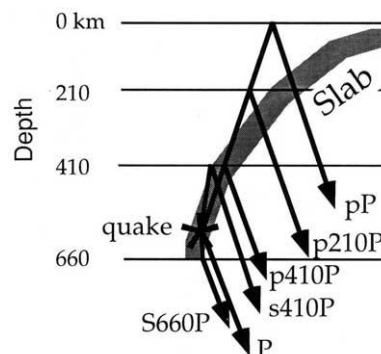


FIG. 2 Typical geometry of an earthquake and the structures that we observe. The shaded region marks the trend of the earthquakes that define the Wadati-Benioff zone.

the character of the earthquake. The traces are shifted in time so that the picked phase arrives at the same time for all traces, normalized to a peak amplitude of unity and searched for plane wave arrivals.

We do not combine earthquakes because the mechanisms and rupture time histories vary, and also we do not assume that each earthquake will illuminate the same structures. We do not cross-correlate to isolate secondary arrivals, as some of the arrivals we seek initially travel upwards, others downwards; some energy emerges from the source region as an S wave, some as a P wave. Only for cases in which the source is small compared to the seismic wavelengths is cross-correlation or deconvolution effective^{2,6}.

The earthquake depth is required to estimate the depth of the reflectors imaged here. The uncertainty in earthquake depth is least for the four events relocated with both pP and P arrivals as part of an inversion for three-dimensional slab structures⁸. The difference in earthquake depth between locations in a one-dimensional and a three-dimensional velocity model is ~1 km (ref. 8). The uncertainty in earthquake depth is difficult to state precisely, but is probably less than 10 km.

The depth uncertainty is greater for the two events taken from the International Seismic Centre (ISC) catalogue. These depths are probably accurate to ± 15 km, depending mostly on whether the correct pP phases were chosen and on the accuracy of the assumed velocity structure above the earthquake⁹. For example, ISC depth estimates using the pP–P times are 5 km deeper on average than depths determined using P times alone, and differences between the two depths have a variance of 11 km. The errors in the two measurements are largely independent, as errors in P locations mostly depend on velocity anomalies beneath the earthquakes whereas errors in pP–P locations mainly depend on velocity anomalies above the earthquakes.

Figure 3a and b compares data and synthetic seismograms for 70 of the 279 waveforms analysed from the 3 April 1985 event, which is near 476 km depth in the Bonin subduction zone. It is clear that the signals we seek are small compared with the P-wave coda and that we must stack many records to resolve the weak reflections expected from the upper-mantle discontinuities.

A slant stack, also called a radon transform by mathematicians and a p – τ stack in the oil industry, resolves plane-wave arrivals while minimizing incoherent signals. A point in the slant stack represents the sum along a line through a seismic section such as those shown in Fig. 3a and b. Each line is described by its slowness (reciprocal of apparent velocity) across the array and its delay time when the line intersects a reference station. Figure 3c and d compares the stacks at the P-wave slowness for the data and the synthetic seismograms. Figure 3e and f compares the stacks at the pP slowness. The stacking process has reduced the noise enough that we can see a minor arrival 20 s after the

TABLE 1 Earthquakes

Date	Region	Depth (km)	Latitude (N)	Longitude (E)	Mag	'660' depth	'410' depth	'210' depth
28 April 1981	Tonga	540*	-23.7°	180.0°	6.0	685	387	208
4 January 1982	Mariana	606*	18.0°	145.6°	6.1	—	408	214
3 April 1985	Bonin	476*	28.2°	139.5°	6.0	677	—	207
16 June 1986	Tonga	559*	-22.0°	-178.9°	6.3	690	398	202
27 October 1987	South America	604†	-28.7°	-62.9°	6.0	686‡	406	220
3 May 1991	Bonin	462†	28.0°	139.6°	6.0	686	—	215

* Depth from relocation and velocity inversion⁶.

† Depth from pP–P location (PDE Bulletin).

‡ Depth from 2-s rather than 4-s period slant stack.

P wave, but only in the stack at the P-wave slowness. This arrival, as well as the energy 60 s after P, will be interpreted below. The pP arrival is noticeably more complicated than the P arrival in the data, probably because of the water layer and structural complexity where pP reflects from the surface of the Earth.

Figure 4 shows the logarithm of the envelope of the slant stacks of records from this earthquake. The envelope is the amplitude of the analytical slant-stack time series and serves only to eliminate the four or five zero crossings in each arrival apparent in Fig. 3c and e. The use of the logarithm with colour allows a factor of 100 variation in amplitude to be presented.

Many arrivals can be seen in Fig. 4. Note that the window between P and pP is very clean, so that signals with amplitudes from 2 to 10% of the direct P wave can be identified. The arrivals are unlikely to be aftershocks because most arrive with slownesses different from the main shock. The similarity of the sequences of arrivals following earthquakes in similar locations also argues against their interpretation as aftershocks.

The signal-to-noise levels are similar for slant-stacks of period 1, 2 and 4 s. It seems that the improved instrument response at shorter periods is offset by larger-amplitude scattered waves and less coherence across the array. The stacks for 4-s period show most clearly the signals from the '410' and '660' reflectors, as well as a signal reflecting near 210 km depth.

Identification of mantle reflections

Four earthquakes show the arrival that travels out from the source as an S wave and converts to a P wave near 660 km depth. A fifth event, indicated in Table 1, only shows a clear arrival at higher frequency. Figure 5 shows the slant stacks of the P wave and the following 30 s for the four events. The traces are aligned so that an S-to-P conversion at 660 km depth would appear centred on the purple vertical line. An arrival appears for each event ~2–3 s later than expected. The amplitude of this arrival is ~10% of the direct P wave, in accord with the prediction of synthetic seismograms.

Arrival times of these S-to-P conversions correspond to reflector depths of 677, 685, 686, 686 and 690 km (Table 1). These depths were calculated by solving for the location of a flat reflector that produces an S-to-P converted arrival with the

observed delay time after the direct P wave. Other possible origins include P-to-P reflections from a horizon that varies between 340 and 500 km depth, S-to-P reflections from a horizon that varies between 380 and 520 km depth, or reflections from dipping surfaces. The consistency of depths, the prediction of such arrivals in reflectivity synthetics and the observation that these arrivals generally show slightly steeper angles of incidence than the direct P wave, as expected for S-to-P conversions in transmission, rather than the slightly less steep angles of incidence expected for reflections and conversions above the source, indicate that the first explanation is correct. The uncertainty in these boundary depth estimates would be less than 10 km, mainly due to the possibility of some dip in the boundary, if the exact depth of the earthquake were precisely known.

The events in Tonga produce '660' depths of 685 and 690. Earlier work on earthquakes with similar focal depths suggested that this boundary was roughly 10 km shallower¹¹. Our earthquake depths, based on pP–P measurements and three-dimensional upper- and lower-mantle velocity structure⁸, are more accurate. The previous work¹¹ measures arrivals that contain energy at <1 s period, whereas we use arrivals of 4 s period. Their arrivals convert nearer the deep seismicity than the arrivals that we examine, as the array that they examine is in Australia whereas ours is in North America. Their array aperture is ~10 km; ours is 1,500 km, allowing greater slowness resolution, although their study has many more events at various depths. We stack seismograms for one event at a time; they stack envelopes for multiple events, so our study is more sensitive to structure that varies laterally. Given these differences, it is not surprising that slightly different depths for the '660' are found.

The slant stacks for all six events are shown in Fig. 6. The primary arrivals, P and pP, appear red and yellow; they have high amplitude and are well defined. The purple triangles above the upper four stacks locate the S'660'P arrivals that were shown in more detail in Fig. 5. For four events, an arrival that we identify as p'410'P is marked by the blue triangles. It may also be present for the other two events, but would be obscured by S'660'P. This reflector, measured with the assumption of horizontal layering, lies at depths of 387, 398, 406 and 408 km for the four events. If these reflections were S-to-P conversions

FIG. 3 a, Seismograms from 70 of the 279 stations used to study the 3 April 1985 earthquake, which cover only a quarter of the aperture of our array. All seismograms are aligned on the pP phase. b, Reflectivity seismograms for the same 70 stations. The IASP91 velocity model is used³¹, and the centroid moment tensor mechanism³² is assumed. The synthetics are filtered in the same way as the data. c, d, Stacks of the data and reflectivity seismograms with the P arrivals aligned. e, f, Stacks of the data and reflectivity seismograms with the pP arrivals aligned.

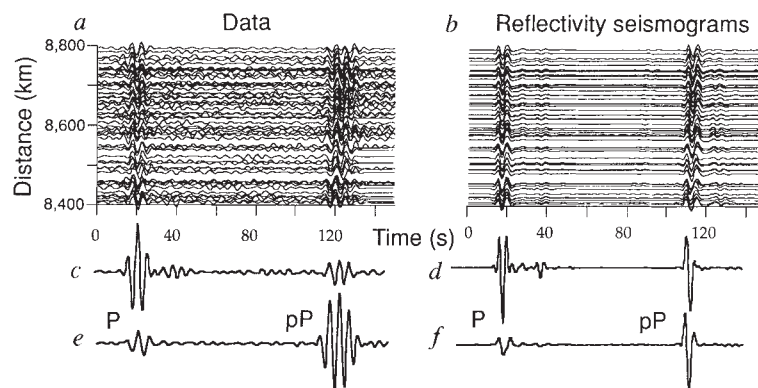
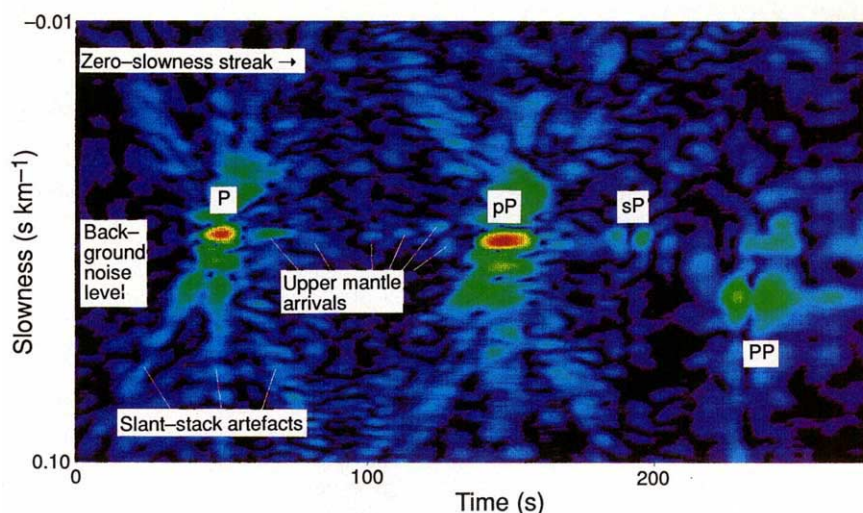


FIG. 4 Slant stack of 3 April 1985 Bonin earthquake. Seismograms from 279 stations are aligned on the P wave and slant-stacked in the time window between P and pP. The logarithms of the envelopes of the stacks are shown. 'Hotter' colour indicates more energy in the stack, with red representing 100 times larger amplitude than black. For example, pP arrives ~ 100 s after P and differs in slowness across the array by 0.003 s km^{-1} . This difference in slowness corresponds to a difference in angle of incidence of $\sim 1^\circ$, or a difference of 5 s between the pP–P times at the opposite ends of the array.



from horizontal reflectors, they would originate at depths greater than 900 km; if they were underside S-to-P conversions, they would originate at depths of 450–500 km. Because little structure is expected at these depths and the amplitude of the reflections is similar to the amplitude of the P-to-P reflections expected from 410 km in the mantle velocity structure IASP91, it is likely that we are observing reflections from the '410'.

If the earthquakes are actually shallower than we estimate, making the depth to the deeper discontinuity more normal (closer to 660 km), then the depth to the upper discontinuity becomes more anomalous (further from 415 km). If the earthquakes are actually deeper, the deeper boundary then behaves more anomalously.

The prominent arrival ~ 50 s before pP, identified by the red triangles, is clear for all of the events. The constancy of the time difference from pP suggests that this arrival is a P reflection rather than an S-to-P conversion. The depths estimated from the time difference between this arrival and pP range from 202 to 215 km. No consistent S-to-P conversion from this boundary, which would appear between p210P and pP, is seen. Perhaps the shear wave is attenuated, or the P velocity contrast is greater,

or possibly the conversion is masked by other reflections from above 210 km depth. Several arrivals are less than 50 s before pP, and are probably due to shallow reflectors, but the arrivals appear with a range of slownesses and are not consistent between events.

In all cases except for the 1987 event, the most prominent arrivals about 50 s before pP are between the P and pP in slowness. From this observation, the reflector probably dips less than 20° from horizontal. The reflection coefficient is 3–10%. This amplitude argues against scattering, which would require a large velocity contrast and strong focusing to reflect so much energy so consistently for earthquakes 250–500 km distant from the scatterer. The presence of reflected P waves of 4-s period,

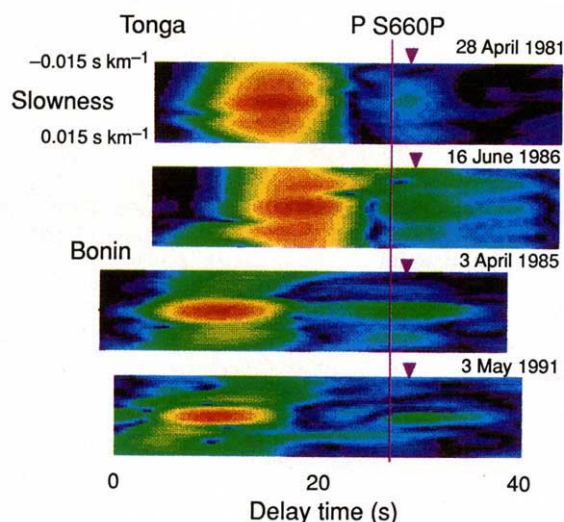


FIG. 5 Slant stack of the shallowest four of the six events used in this study. The stacks are aligned so that the S-to-P conversion from a boundary at 660 km depth would appear at the time of the vertical purple line. Arrivals are apparent at times corresponding to depths of 677, 685, 686 and 690 km. Slowness is measured relative to the P wave, which is assigned a slowness of 0.0 s km^{-1} .

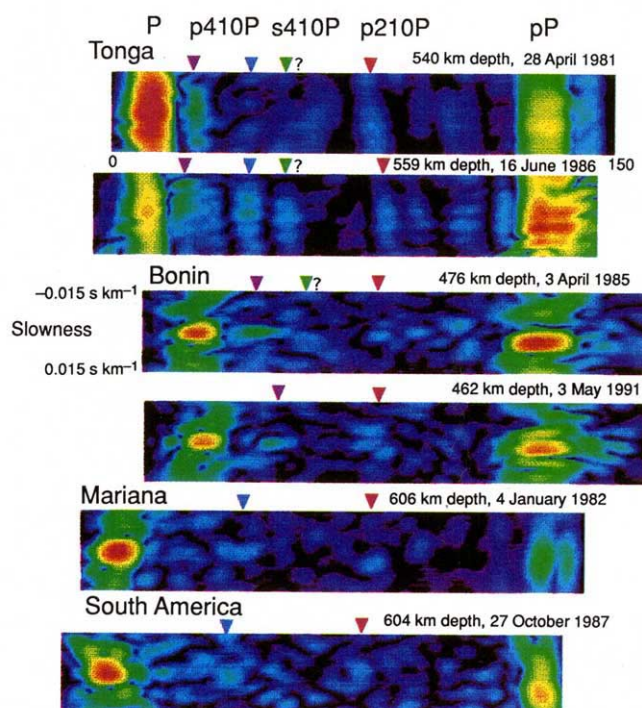


FIG. 6 Slant stacks of all six events studied. The stacks are aligned so that the phase pP arrives at the same time. Four arrivals are apparent at times indicating depths near 410 km (blue triangles) and six arrivals from near 210 km (red triangles; Table 1). Purple triangles mark arrivals interpreted as S'660P in Fig. 5, and green shows possible s'410P arrivals. Each slant stack shows 150 s.

which have 30 km wavelength, requires that the reflector be less than a quarter wavelength (8 km) thick.

Cause of reflectors at 210, 400 and 680 km

The five observations of upper-mantle seismic discontinuities near 660 km depth suggest that the γ -phase olivine to perovskite plus magnesio-wüstite boundary is depressed by 20–30 km compared to global averages based on long-wavelength measurements^{2,5,6}. This variation is somewhat greater than the 20-km depression observed to the northwest of subduction in the north-west Pacific with long-period shear waves², whose resolution is $\sim 1,000$ km. The resolution of our high-frequency observations is an order of magnitude finer because we are using waves five times higher in frequency and because the conversion point is near the source rather than near the midpoint of the ray path. We sample the '660' about 100 km from the intersection of the seismicity that defines the Wadati-Benioff zone and 660 km depth, toward the direction from which the subducting lithosphere is approaching (the right, in Fig. 2). The '660' boundary in this direction has been estimated to lie at a more normal depth than on the other side in the Kurils². We therefore observe more depth variation than is seen with longer-wavelength probes. Depression of the boundary by 25 km suggests a decrease in temperature of ~ 300 °C ($\pm 30\%$)¹¹. And this elevation is measured outside the region where simple models of through-going subducting slabs place the core of the slab¹², requiring that the thermal anomaly be spread by broadening of the slab rather than by thermal conduction alone.

The four observations of upper-mantle seismic discontinuities near 410 km depth suggest that the α -to- β phase boundary is elevated by ~ 15 km near the core of subducting slabs. Our reflection points range from the middle of the seismicity at 410 km depth to ~ 100 km to the side of the seismicity. As our observations are made in areas above deep earthquakes, we are measuring the depth of the phase transition either in material that is subducting or that has been next to subduction for a long time. An elevation of the equilibrium $\alpha \rightleftharpoons \beta$ phase boundary of 15 km would result from 400 °C ($\pm 30\%$) difference in temperature between subduction zones and normal mantle^{13,14}. This is less than the expected thermal anomaly in the core of the slab. But as the wavelength of features that can be resolved in this study is 50–100 km, we still may not be observing the full elevation of the discontinuity if the coldest zone is narrow.

Mineral physics experiments show that the $\alpha \rightleftharpoons \beta$ phase transition is kinetically hindered at temperatures below 520–900 °C, at least on laboratory timescales¹⁵. Seismic studies have suggested that the depth of this transition is depressed near slabs¹⁶, is elevated^{17,18} or has no measurable deflection⁵, and a small anticorrelation between the depth to the 660 and 410 km discontinuities has been observed⁶. Our observations of a boundary elevated by 15 km suggest that the phase transition can proceed at temperatures 400 °C colder than ambient mantle¹³. It is still possible that isolated pockets of olivine remain untransformed, or that the cold centre of the slab forms a thin, untransformed tongue of metastable olivine.

Recent work indicates that this transition may initiate and lubricate deep earthquakes, overcoming the high friction theoretically expected to prevent seismic faulting^{19,20}. Our results indicate that the transformation progresses at temperatures several hundred degrees colder than the normal mantle geotherm. If metastability is to account for deep earthquakes, it must persist to near 660 km depth, despite the adiabatic temperature increase of ~ 200 °C from 410 to 660 km depth¹³. Alternatively, some compositional or grain-size influence on metastability may allow the penetration of the α phase to the depths of the deepest earthquakes, but this seems unlikely.

Published observations and interpretations of a seismic velocity discontinuity near 210 km depth span the past 30 years^{5,7,21,22}. Long-period, wide-angle seismic profiles have often been interpreted to show structures near 210 km depth. A velocity increase

near 200 km depth underlain by a minor velocity decrease is seen beneath Australia²³, but the sharpness is difficult to resolve. A velocity decrease near 160 km depth, underlain by an increase in velocity near 190 km, is inferred beneath northwestern Eurasia²⁴. These two studies examined areas beneath stable continents. A large velocity contrast underlies Japan, somewhat shallower than 200 km depth²⁵. Under Tibet, a reflector is observed near 200 km depth²⁶.

Reflections from near 210 km depth are not seen by recent long-period shear-wave studies^{5,27} under oceans; the present study, however, has the advantage of a more favourable signal-to-noise ratio, more localized sampling and more sensitivity to moderately dipping reflectors. Another explanation of this discrepancy would be that the reflector is laminated, with little long-period signature. Perhaps the feature is confined to a small area around subduction zones. But the volume of literature reporting seismic velocity changes near 210 km depth suggests that a global boundary may be present.

The nature of the reflector required by Fig. 6 is not simple to deduce. A depth of 210 km has been proposed as the level of a compositional boundary^{22,28}. The constancy of depth adjacent to subduction zones, where material passes through the 210-km level, is difficult to reconcile with a compositional boundary. This depth has also been proposed as the bottom of partial melting, which would correspond to the base of the low-velocity zone, and this is the most likely interpretation. A discontinuity near this depth has been proposed as the base of the lid beneath continents, but it seems to lie deeper and mainly beneath older continents²⁷, whereas we observe the structure beneath oceans.

The reflections from the '410' show that the $\alpha \rightarrow \beta$ boundary is elevated by ~ 15 km within cold, subducting material, consistent with a temperature anomaly of ~ 400 °C. This amount of relief on the boundary argues against penetration of metastable olivine to the depth of the deepest earthquakes. The 25-km depression of the boundary at which the γ -phase of olivine transforms to perovskite plus magnesio-wüstite is greater than that observed at longer wavelength, and consistent with temperatures 300 °C colder than normal. Such temperature contrasts are about twice as large as those inferred from large-scale global variations of discontinuity depth. Measurements of long-period shear waves^{3,6} find ~ 15 km of depth variation (zero to peak) for the '660' and 5–10 km for the '410'. These studies indicate lateral variations of ~ 100 °C. Lateral compressional-wave velocity variations have been interpreted to indicate 120 (± 100) °C temperature variations in the mid-mantle²⁸.

Temperature variations of 300–400 °C are consistent with thermal models of slabs. The core of the slab may be ~ 800 °C colder than ambient mantle¹², so a contrast of 400 °C at 410 km depth is the result one would expect from this fairly long-wavelength probe. Near the '660', it is more difficult to know what temperature contrast to expect ~ 100 km towards the direction of approach of the subducting plate. The simplest, narrowest slab models^{12,20} predict only a small temperature perturbation at our S'660'P conversion point. The models in which viscosity increases by more than a factor of 10 below the '660' (ref. 30) and the models in which the anomalously cold zone is broad¹² do predict a few hundred degrees of anomaly, as we observe. $\square$

Received 30 December 1991; accepted 24 March 1992.

1. Jeanloz, R. *Geophys. Res. Lett.* **18**, 1743–1746 (1991).
2. Shearer, P. M. & Masters, T. G. *Nature* **355**, 791–795 (1992).
3. Silver, P. G., Carlson, R. W. & Olson, P. A. *Rev. Earth planet. Sci.* **16**, 477–541 (1990).
4. Helmberger, D. V. & Engen, G. *J. geophys. Res.* **79**, 4017–4029 (1974).
5. Shearer, P. M. *J. geophys. Res.* **96**, 18147–18182 (1991).
6. Revenaugh, J. & Jordan, T. H. *J. geophys. Res.* **96**, 19763–19780 (1991).
7. Wald, L. A. & Heaton, T. H. *J. geophys. Res.* **96**, 12099–12125 (1991).
8. Van der Hilst, R., Engdahl, E. R., Spakman, V. & Niolet, G. *Nature* **353**, 37–43 (1991).
9. Stark, P. B. & Frohlich, C. *J. geophys. Res.* **90**, 1859–1869 (1985).
10. Richards, M. A. & Wicks, C. W. Jr *Geophys. J.* **101**, 1–35 (1990).
11. Ito, E. & Takahashi, E. *J. geophys. Res.* **94**, 10637 (1989).
12. Fischer, K., Creager, K. & Jordan, T. H. *J. geophys. Res.* **93**, 4773 (1988).
13. Akaogi, M., Ito, E. & Navrotsky, A. *J. geophys. Res.* **94**, 15671–15688 (1989).
14. Katsura, T. & Ito, E. *J. geophys. Res.* **94**, 15663–15671 (1989).

15. Sung, C. M. & Burns, R. G. *Earth planet. Sci. Lett.* **32**, 165–170 (1976).
16. Iidaka, T. & Suetetsu, D. *Nature* (in the press).
17. Solomon, S. C. & U, K.T.P. *Phys. Earth planet. Sci. Lett.* **11**, 97–108 (1975).
18. Helfrich, G. R., Stein, S. & Wood, B. J. *J. geophys. Res.* **94**, 753–763 (1989).
19. Green, H. W. & Burnley, P. C. *Nature* **341**, 733–737 (1989).
20. Kirby, S. H., Durham, W. B. & Stern, L. A. *Science* **252**, 216–225 (1991).
21. Lehmann, I. *Ann. Geophys.* **15**, 93–118 (1959).
22. Anderson, D. L. *J. geophys. Res.* **84**, 7555–7561 (1979).
23. Bowman, J. R. & Kennett, B. L. N. *Geophys. J. Int.* **101**, 355–366 (1990).
24. Given, J. W. & Helmberger, D. V. *J. geophys. Res.* **85**, 7183–7184 (1980).
25. Kanamori, H. *Bull. Earthquake Res. Inst.* **45**, 657 (1967).
26. Kato, M. & Hirahara, K. *Geophys. J. Int.* **106**, 551–558 (1991).
27. Revenaugh, J. & Jordan, T. H. *J. geophys. Res.* **96**, 19781–19810 (1991).
28. Anderson, D. L. & Bass, J. *Nature* **320**, 321–328 (1986).
29. Gurnis, M. & Hager, B. H. *Nature* **335**, 317–321 (1988).
30. Duffy, T. & Ahrens, T. J. *J. geophys. Res.* (in the press).
31. Kennett, B. L. N. *IASPI 1991 Seismological Tables* (Research School of Earth Sciences, Australia National Univ., 1991).
32. Dziewonski, A. M. & Woodhouse, J. H. *J. geophys. Res.* **88**, 3247–3271 (1983).

ACKNOWLEDGEMENTS. We thank Q. Williams, T. Lay, H. Houston, E. Knittle and W. Ellsworth for discussions, P. Shearer and E. R. Engdahl for helpful reviews, the Lamont-Doherty Geological Observatory, J. Evans, B. Julian, A. Walter and J. Luetgert for software, R. Somera, J. Mori, R. Benson, S. Nava and D. Depaulo for assistance in recovering data, and E. R. Engdahl and R. Van der Hilst for providing unpublished depths for four earthquakes.

Successive action of DnaK, DnaJ and GroEL along the pathway of chaperone-mediated protein folding

Thomas Langer^{*}, Chi Lu[†], Harrison Echols[‡], John Flanagan[‡], Manajit K. Hayer^{*} & F. Ulrich Hartl^{*,§}

^{*} Program in Cellular Biochemistry and Biophysics, Rockefeller Research Laboratories, Sloan-Kettering Institute, 1275 York Avenue, Box 520, New York, New York 10021, USA

[†] Department of Molecular and Cell Biology, University of California, Berkeley, California 94720, USA

[‡] Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, Connecticut 06511, USA

The main stress proteins of *Escherichia coli* function in an ordered protein-folding reaction. DnaK (heat-shock protein 70) recognizes the folding polypeptide as an extended chain and cooperates with DnaJ in stabilizing an intermediate conformational state lacking ordered tertiary structure. Dependent on GrpE and ATP hydrolysis, the protein is then transferred to GroEL (heat-shock protein 60) which acts catalytically in the production of the native state. This sequential mechanism of chaperone action may represent an important pathway for the folding of newly synthesized polypeptides.

PROTEINS can fold *in vitro* without the help of additional components, demonstrating that the amino-acid sequence contains the full information to specify the native conformation^{1,2}. It was therefore assumed that the folding of proteins *in vivo* occurred by an essentially spontaneous process as well. But then it was discovered that the folding of certain newly synthesized proteins depends on so-called 'molecular chaperones'³, or 'polypeptide chain binding proteins'⁴. These proteins have the ability to interact with many unfolded or partially denatured proteins apparently without specific recognition of defined sequence motifs. The involvement of molecular chaperones in *de novo* protein folding has been established for the constitutively expressed stress proteins of the heat-shock proteins Hsp70 and Hsp60 families^{5–8}. Defining the functions of these components will be crucial for understanding the mechanisms of protein folding within cells.

Although structurally unrelated, members of both classes of molecular chaperones use the energy of ATP hydrolysis to release bound substrate proteins. The Hsp70s probably act as monomers or dimers, whereas the Hsp60s (the so-called

chaperonins⁹) are functional as high-molecular-weight oligomeric structures that are composed of two stacked seven-subunit rings. Members of the Hsp70 family are involved in diverse functions, including maintaining proteins in an unfolded state^{10–13}, dissociating protein aggregates, and facilitating renaturation^{5,14,15}. The Hsp70 homologue of *E. coli*, DnaK, cooperates functionally with two additional heat-shock proteins, DnaJ and GrpE (refs 16–18). Members of the Hsp60 family *in vivo* and *in vitro* mediate the folding of proteins to the native structure and facilitate oligomeric protein assembly^{19–29}. Folding may occur by a process of stepwise ATP-dependent release from the oligomeric GroEL structure²⁸.

Studies on the cellular functions of heat-shock proteins have focused mainly on one class of components, so we do not yet have an integrated view of the roles of the different chaperone families in protein folding. Hsp70s and Hsp60s seem to interact with different structural elements of unfolded polypeptides. Consistent with their binding to nascent chains³⁰, members of the Hsp70 family interact with synthetic peptides³¹ and with polypeptides in extended conformations³². In contrast, the Hsp60 of *E. coli*, GroEL, may recognize secondary structure elements³³ and stabilize bound proteins as conformational intermediates formed early in folding, like the 'molten globule'²⁸. That division of labour between Hsp70s and Hsp60s is physiologically significant is shown by the folding of proteins in mitochondria. During import from the cytosol, precursor proteins traverse the mitochondrial membranes as extended chains³⁴. The newly translocated polypeptides first interact with the DnaK-related Hsp70 in the mitochondrial matrix^{35–37}. For a number of proteins this interaction is necessary but not sufficient for folding and assembly, which depend on the subsequent transfer of the polypeptide chains to Hsp60 (refs 19, 21, 37), the mitochondrial GroEL homologue³⁸. A DnaJ-like activity may participate in the process, given that a DnaJ-related protein has recently been identified that is possibly localized in mitochondria³⁹.

As the chaperone proteins of mitochondria have homologous equivalents in the *E. coli* cytosol, it may be that the pathways of chaperone-mediated protein folding in both compartments

§ To whom correspondence should be addressed.

are similar. Model substrate proteins were used to establish the minimal number of heat-shock proteins of *E. coli* necessary to reconstitute this protein-folding reaction. We found that DnaK, DnaJ and GroEL interact sequentially with the folding polypeptide in a reaction coupled by GrpE. GroEL in substoichiometric concentrations can mediate the folding of multiple substrate molecules of a pool of DnaK/DnaJ-stabilized chains. We propose that DnaK and DnaJ act to prevent premature misfolding and aggregation, after which GroEL mediates proper folding. In this model, the chaperone proteins have distinct roles in the structural maturation of newly synthesized proteins.

Cooperation of DnaK and DnaJ

Mitochondrial Hsp70 binds the unfolded polypeptide chains that are newly translocated into the organelles^{35,36}. We attempted to reproduce this type of interaction using the Hsp70 homologue, DnaK, purified from *E. coli*. Rhodanese from bovine mitochondria⁴⁰ was used as the substrate protein which was unfolded in 6 M guanidinium-HCl. In spite of being monomeric, denatured rhodanese will not refold efficiently when diluted from denaturant into buffer alone. The protein aggregates at sub-micromolar concentrations⁴¹. DnaK suppressed this aggregation, but only when present at a high molar excess (Fig. 1a). The association of DnaK with an intermediate conformation of rhodanese, formed rapidly on dilution from denaturant^{42,43}, might therefore require an additional chaperone component. We analysed the effect of the heat-shock protein DnaJ, which interacts functionally with DnaK¹⁶⁻¹⁸. Surprisingly, DnaJ itself retarded the aggregation of the denatured protein and even completely prevented it at a fivefold molar excess (Fig. 1a). The stability of the complex was probed using casein, a soluble protein that exposes hydrophobic surfaces under native conditions^{44,45}. We found that casein displaced rhodanese from DnaJ, resulting in aggregation, when it was added to the preformed rhodanese/DnaJ complex (Fig. 1b; compare lanes 1 and 2). Interestingly, rhodanese bound more tightly in the presence of ATP (independently of Mg²⁺ ions) and higher concentrations of the competitor were necessary to displace it (lane 3). DnaJ may indeed bind ATP (J.F., R. Belin and D. Engelman, unpublished observations), although it apparently does not hydrolyse ATP⁴⁶.

We tested the possibility of a synergistic effect of DnaK and DnaJ in preventing rhodanese aggregation. A substoichiometric concentration of DnaJ (versus rhodanese), which did not measurably suppress aggregation, markedly increased the potency of DnaK to interact with the denatured protein (Fig. 1c). Rhodanese was then bound more tightly than by DnaJ alone, as indicated by the reduced ability of casein to displace rhodanese (Fig. 1b; compare lanes 2 and 5). Stability was optimal only in the presence of both ATP and Mg²⁺, thus allowing for ATP hydrolysis (lane 7). In contrast, the effect of high concentrations of DnaK alone to stabilize rhodanese (Fig. 1a) was lost on addition of Mg-ATP and rhodanese aggregated with a half-time of about 20 min (not shown). ATP hydrolysis causes the release of DnaK-bound substrate proteins⁶. Apparently, the way in which DnaK interacts with the folding protein was changed markedly by the simultaneous presence of DnaJ.

Complex formation between rhodanese and chaperone proteins was analysed by gel chromatography. Unfolded rhodanese was added to DnaK and DnaJ at a molar ratio of 1:5:2, respectively, which prevents aggregation (Fig. 1a). Under these conditions, about 80% of rhodanese fractionated on Sephacryl columns as a complex of $M_r \sim 230K$ (Fig. 2A, c). Interestingly, this is the size recently reported for a cytosolic protein complex containing Hsp70, in which mitochondrial precursor proteins are preserved in a competent state for membrane translocation^{11,47}. The interaction between rhodanese, DnaK and DnaJ was not disrupted when Mg-ATP was present throughout the fractionation. In contrast, only 45% of total denatured rhodanese

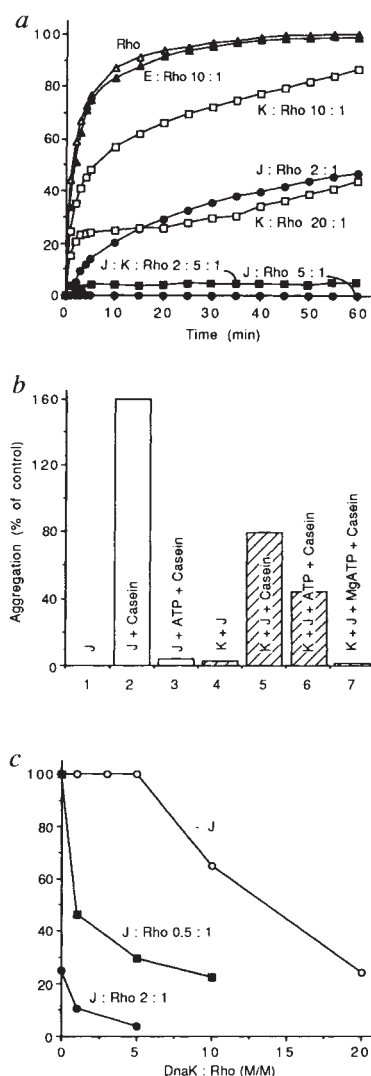
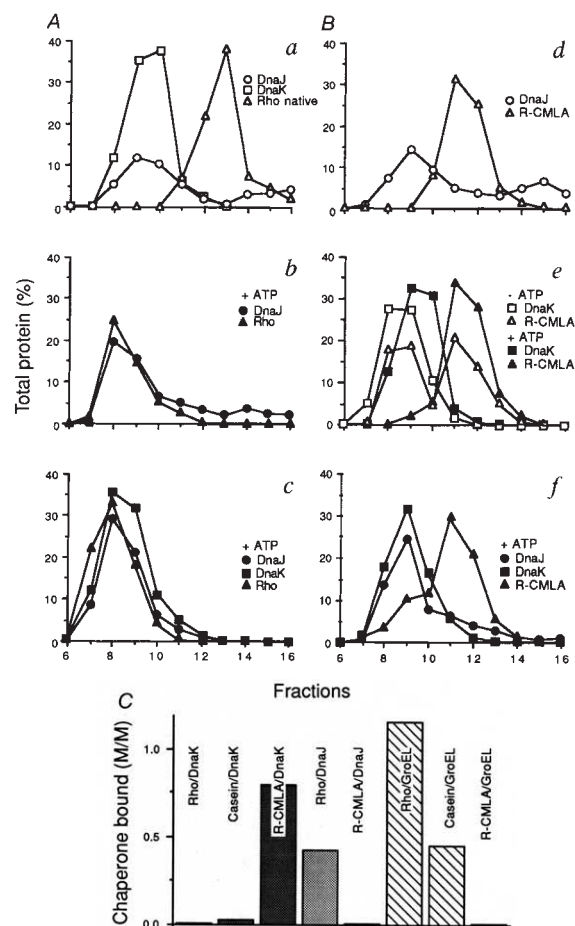


FIG. 1 Cooperation of DnaK and DnaJ in preventing rhodanese (Rho) aggregation. **a**, Time course of aggregation in the absence of added components (Δ), in the presence of DnaK (K) ($\square$), DnaJ (J) ($\bullet$), GrpE (E) ($\blacktriangle$) or DnaK and DnaJ in combination ($\blacksquare$). **b**, Displacement of rhodanese from DnaJ (J) or DnaK/DnaJ by casein. **c**, Suppression of aggregation by DnaK in the absence ($\circ$) and in the presence of DnaJ added at a molar ratio to rhodanese of 0.5:1 ($\blacksquare$) and 2:1 ($\bullet$), respectively.

METHODS. DnaJ, GrpE⁴⁶ and GroE²⁸ were purified from overproducing strains of *E. coli* to $\geq 92\%$ purity and were essentially free of the other chaperone proteins. DnaK⁶⁵ was overproduced under control of the T7 promoter. Protein concentrations were adjusted using the following absorbance values at 280 nm for 0.1% solutions based on quantitative amino-acid analysis and colorimetric protein determination⁶⁶: DnaJ, 0.31; DnaK, 0.20; GrpE, 0.07; GroEL, 0.20; GroES, 0.15. Rhodanese from bovine liver ($\sim 95\%$ pure; Sigma) was denatured at $46\mu M$ in buffer A (6 M guanidinium-HCl, 30 mM Tris-HCl, pH 7.4, 5 mM dithiothreitol (DTT)) by incubation for 60 min at room temperature. **a**, Unfolded rhodanese was diluted 100-fold ($0.46\mu M$ final concentration) into buffer B (10 mM MOPS-KOH, pH 7.2, 50 mM KCl) at $25^\circ C$ containing DnaK at concentrations of $0\mu M$, $4.6\mu M$ and $9.2\mu M$ or DnaJ at $0.92\mu M$ and $2.30\mu M$, GrpE at $4.6\mu M$, and DnaK/DnaJ at $2.30\mu M/0.92\mu M$, respectively. Aggregation (turbidity) was measured at $320 nm$ ⁶⁷. **b**, Rhodanese-chaperone complexes were generated in buffer B containing DnaJ ($2.30\mu M$) or DnaK/DnaJ ($2.30\mu M/0.92\mu M$). Aggregation was measured 20 min after addition of $2.0\mu M$ α_{s1} -casein (Sigma). 1 mM ATP and 5 mM magnesium acetate were present when indicated. Aggregated rhodanese was sedimentable (15 min at $30,000g$). Aggregation in the presence of casein was greater than in control reactions (that is, dilution of denatured rhodanese into buffer), whereas $>90\%$ of casein and $>95\%$ of DnaJ and DnaK remained in the supernatant. **c**, Denatured rhodanese was diluted into buffer B as above, containing $0\mu M$, $0.23\mu M$ and $0.92\mu M$ of DnaJ in the presence of $0-9.2\mu M$ of DnaK. Aggregation was measured after 15 min at $25^\circ C$.

FIG. 2 Gel chromatography of substrate-chaperone complexes. **A, a**, Elution profiles of pure proteins: native rhodanese (Δ), DnaJ ($\square$), DnaK ($\circ$); **b**, complex formation of denatured rhodanese ($\blacktriangle$) with DnaJ ($\bullet$), and **c**, with DnaJ/DnaK ($\bullet$, $\blacksquare$). **B, d**, Gel chromatography of R-CMLA (Δ) incubated with DnaJ ($\circ$); **e**, complex formation of R-CMLA (Δ , $\blacktriangle$) with DnaK ($\square$, $\blacksquare$) in the absence (Δ , $\square$) or presence ($\blacktriangle$, $\blacksquare$) of Mg-ATP, and **f**, with DnaK/DnaJ ($\blacksquare$, $\bullet$) in the presence of Mg-ATP. Protein amounts are given in per cent of total applied to the columns. **C**, Association of various substrate proteins with DnaK, DnaJ and GroEL as indicated, based on gel chromatography of the respective complexes.

METHODS. **A**, Denatured rhodanese was diluted to 1.2 μ M into buffer B containing 4.8 μ M DnaJ (**b**) or 6 μ M/2.4 μ M DnaK/DnaJ (**c**). When indicated, 10 mM magnesium acetate and 5 mM ATP were added. The reactions were incubated for 10 min at 25 $^{\circ}$ C. Ten per cent of each reaction was removed for determination of recoveries. The remainder was applied to a 0.5 $\times$ 10 cm Sephacryl-S200 column (Pharmacia) equilibrated with buffer B or with buffer B containing Mg-ATP. Fractions of 120 μ l were collected. Trichloroacetic acid precipitates were analysed by SDS-PAGE and protein amounts on Coomassie-stained gels were quantified by laser densitometry. In parallel experiments DnaJ, DnaK, native rhodanese and denatured rhodanese diluted into buffer B were fractionated under identical conditions (**a**). Aggregates of rhodanese were not eluted from the column by buffer B but were recovered by a wash with 8 M urea (not shown). Sizes of rhodanese/DnaJ and rhodanese/DnaK/DnaJ complexes were determined on 1 $\times$ 20-cm S300 columns. **B**, 4.4 μ M of fully reduced, carboxymethylated bovine α -lactalbumin (R-CMLA) was incubated with DnaJ (4.4 μ M), DnaK (4.4 μ M), or DnaJ/DnaK (4.4 μ M/4.4 μ M) in buffer B for 10 min at 25 $^{\circ}$ C (for DnaJ) or at 37 $^{\circ}$ C (when DnaK was present). Binding of DnaK was more efficient at 37 $^{\circ}$ C; binding of DnaJ was the same at 37 $^{\circ}$ C and 25 $^{\circ}$ C. When indicated, Mg-ATP was added before further incubation for 5 min at 37 $^{\circ}$ C. Complex formation was analysed as in **A**. **C**, DnaK (4.5 μ M), DnaJ (4.5 μ M) and GroEL-14mer (1 μ M) were incubated for 10 min at 25 $^{\circ}$ C (37 $^{\circ}$ C in the case of DnaK) with a 4-fold molar excess of rhodanese (diluted from 6 M guanidinium hydrochloride), casein or R-CMLA as indicated. Binding of rhodanese to DnaK was unchanged at 25 $^{\circ}$ C. Analysis by gel chromatography was carried out using S200 and S300 columns. The amount of bound ligand was estimated on the basis of the amount of ligand in the peak fractions of the respective chaperone proteins not overlapping with the distribution of the free ligand.



was recovered as a similarly sized complex with DnaJ alone (Fig. 2A, b), even when the concentration of DnaJ was doubled compared with that used in the presence of DnaK. Apparently, the rhodanese/DnaJ complex dissociates fast enough so that some dissociation occurs during gel filtration. This effect was more pronounced for the interaction between DnaK and rhodanese, thus preventing the isolation of a complex between these proteins. Aggregated rhodanese was not eluted from the column with the standard buffer but was recovered by a wash with 8 M urea (not shown). The exact stoichiometry of the rhodanese-DnaJ and rhodanese-DnaK-DnaJ complexes has yet to be determined.

Differential recognition

Is DnaJ generally required for DnaK to interact efficiently with unfolded substrate proteins? Rhodanese diluted from denaturant shares the known property of proteins to undergo collapse to a compact folding intermediate on the millisecond timescale^{48,49}. To mimic the interaction of DnaK with an extended polypeptide chain emerging from a ribosome early in translation, we used reduced carboxy-methylated α -lactalbumin (R-CMLA) as substrate. This model protein maintains an extended conformation without any stable secondary structure in the absence of denaturant, as established by circular dichroism spectroscopy and urea-gradient electrophoresis (J. Ewbank and T. Creighton, manuscript in preparation). DnaK bound efficiently to R-CMLA (Fig. 2B, e). Binding was somewhat less efficient at 25 $^{\circ}$ C than at 37 $^{\circ}$ C, as shown for eukaryotic Hsp70 (ref. 32). In contrast, gel filtration analysis revealed that DnaJ and GroEL had only a very low affinity for the unfolded protein (Fig. 2B, d and C), whether or not ATP was present. But DnaJ and GroEL were able to interact with α -lactalbumin when the native protein was unfolded to the so-called molten-globule state⁵⁰ (unpublished observations). Casein, which appears to

share certain properties of a compact intermediate such as the exposure of hydrophobic surfaces^{44,45}, was also recognized by both DnaJ and GroEL, but not by DnaK (Fig. 2C). DnaJ-bound casein could not be quantified reliably by gel fractionation owing to insufficient separation between free and bound forms. The affinity of DnaJ for casein has been demonstrated by the competition experiments shown in Fig. 1b.

We were interested to see how DnaJ would influence the interaction between DnaK and the extended polypeptide. Addition of DnaJ (with or without ATP) did not increase the affinity of DnaK for R-CMLA, but gel chromatography revealed that in the presence of Mg-ATP, DnaJ (equimolar with DnaK) retarded the release of the unfolded protein from DnaK (Fig. 2B, e and f). Thus, there appears to be a weak three-protein interaction involving DnaJ, DnaK and R-CMLA. The strong interaction observed for DnaJ, DnaK and rhodanese may result from the ability of DnaJ and DnaK together to trap a partially folded polypeptide chain as a folding intermediate (a state not effectively attained by R-CMLA). Evidence for this idea was provided by the spectral analysis described below.

What are the conformational properties of chaperone-stabilized rhodanese? The lack of tryptophan residues in DnaJ⁵¹ allowed us to analyse the intrinsic tryptophan fluorescence of DnaJ-bound rhodanese, thus providing an indicator of its tertiary structure content. Generally, increased exposure of tryptophan residues to solvent on unfolding results in a red shift of tryptophan fluorescence excited at 295 nm. The wavelength of maximum emission of DnaJ-stabilized rhodanese (containing eight tryptophans⁴⁰) was at 343 nm, shifted half-way towards that of the form unfolded by guanidinium-HCl. This indicated that the tryptophan residues were still in a non-native, but already more hydrophobic environment than in the fully unfolded state (Fig. 3a). Shielding of a hydrophobic protein surface by the associated chaperone could contribute to creating

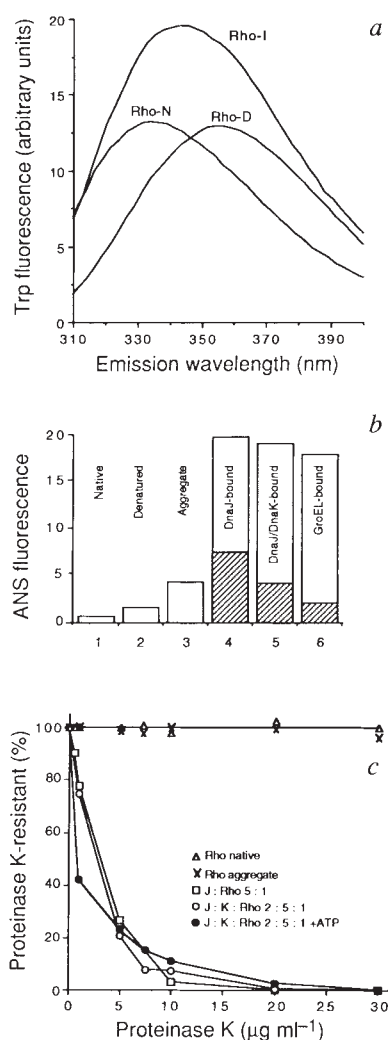


FIG. 3 Conformational properties of chaperone-associated rhodanese. *a*, Trp fluorescence of native (Rho-N), denatured (Rho-D) and of DnaJ-bound forms (Rho-I). *b*, ANS fluorescence. *c*, Intrinsic resistance to proteinase K digestion of native (Δ), aggregated ($\times$), DnaJ-bound ($\square$), and DnaK/DnaJ-bound rhodanese in the absence ($\circ$) or presence ($\bullet$) of Mg-ATP. **METHODS.** Fluorescence spectra were recorded on a SPEX Fluorolog 211. *a*, Trp fluorescence was excited at 295 nm²⁸. Rhodanese/DnaJ complex (0.3 μM /1.0 μM) was generated in the presence of 1 mM ATP by diluting rhodanese from 6 M guanidinium-HCl as in Fig. 1. Rhodanese aggregates (<10% of total rhodanese) were sedimented for 15 min at 30,000g. The rhodanese/DnaJ complex was stable for at least 25 min. Background fluorescence (~25% of total fluorescence) of chemically identical reactions lacking rhodanese was subtracted. Background fluorescence of the rhodanese/DnaK/DnaJ complex (0.3 μM /1.0 μM /0.6 μM) was ~60% of total fluorescence due to the presence of Trp in DnaK (not shown). *b*, Various forms of rhodanese were incubated with a 20-fold molar excess of ANS^{28,49}. Binding to DnaJ (4) and DnaK/DnaJ (5) was as described. Rhodanese/GroEL complex (6) (0.3 μM /0.8 μM) was formed by dilution of denatured rhodanese into GroEL containing solution. Spectra (excitation 390 nm, emission 470 nm) were corrected for background fluorescence in the absence of protein. Fluorescence intensities are given in arbitrary units per mole of protein (of monomer in case of rhodanese, DnaK, DnaJ, and of 14-mer in case of GroEL). Hatched areas in columns 4–6 represent ANS fluorescence of chaperone proteins in the absence of ligand (per mol of total chaperone protein). *c*, Denatured rhodanese was diluted to 0.71 μM into buffer B containing DnaJ (2.84 μM) or DnaK/DnaJ (3.55 μM /1.42 μM). When indicated, incubation in the presence of Mg-ATP was performed for 10 min at 25 °C. Small amounts of aggregated rhodanese were removed by centrifugation (see above). Protease treatment was performed for 10 min at 0 °C and was stopped by adding phenylmethylsulfonylfluoride to 2 mM. Trichloroacetic acid precipitates were analysed by SDS-PAGE and immunoblotting using anti-rhodanese antibodies and the luminescence-based detection system ECL (Amersham)²⁸. Exposures were quantified by laser densitometry.

such an environment. The additional presence of DnaK in the complex appeared to reduce the tryptophan fluorescence intensity of rhodanese by ~30% without changing the wavelength of maximum emission (not shown). This could suggest that rhodanese was in a different conformation when bound to DnaK/DnaJ as compared with the DnaJ-associated species. But the presence of tryptophan in DnaK itself limited the potential significance of this analysis (see legend to Fig. 3).

Intermediate emission of tryptophan fluorescence can be observed with certain collapsed folding states containing a fluctuating hydrophobic core that adsorbs the fluorescent compound 1-anilino-naphthalene-8-sulphonate (ANS)^{42,48,49}. ANS fluorescence may be used as an indirect parameter to infer the presence in such folding intermediates of collapsed, but loosely packed secondary structure^{48,49}. In contrast to the native and fully unfolded states, rhodanese stabilized by DnaJ showed significant ANS fluorescence which was blueshifted to about 470 nm, indicating that ANS was in a hydrophobic environment (Fig. 3*b*, lane 4). Higher ANS adsorption was observed with rhodanese bound to DnaK/DnaJ (lane 5), reaching the fluorescence intensities found with the GroEL-bound, protein²⁸ (lane 6). Interestingly, DnaJ itself showed marked ANS fluorescence, indicating that on a molar basis this chaperone protein exposes a greater hydrophobic surface to solvent than DnaK or the GroEL 14-mer (lanes 4–6, hatched boxes).

Chaperone-bound rhodanese was very sensitive to digestion by proteinase K whether or not Mg-ATP was present (Fig. 3*c*). Protease-stable fragments were not observed. Notably, both native and aggregated rhodanese showed high intrinsic resistance to protease (Fig. 3*c*). In sum, DnaK and DnaJ both cooperate in stabilizing the protein substrate as a folding intermediate which is likely to contain an as-yet-undetermined amount of secondary structure but lacks ordered tertiary structure. Similar properties have been proposed for the GroEL-bound protein²⁸.

GrpE coupled transfer to GroEL

Rhodanese stabilized by DnaK/DnaJ or by DnaK alone did not assume its active conformation on addition of Mg-ATP, although our purified DnaK exhibited normal ATP hydrolytic activity (not shown). We incubated DnaK/DnaJ-associated rhodanese with Mg-ATP and GrpE, an *E. coli* heat-shock protein that strongly stimulates the ATPase activity of DnaK when DnaJ is present⁴⁶. This allowed refolding to occur at a slow rate, approaching 10% of the control activity after 2 hours (Fig. 4*a*) and reaching a final yield of about 30% after 8 hours (not shown). These experiments were carried out at a molar ratio of rhodanese:DnaK:DnaJ of 1:5:2. Varying the relative concentrations of the three heat-shock proteins failed to improve the efficiency of renaturation.

Can folding of DnaK/DnaJ-associated protein be mediated by the *E. coli* homologue of Hsp60, GroEL? This chaperonin recognizes rhodanese that is added from denaturant and, regulated by the co-chaperonin GroES, mediates efficient ATP-dependent folding to the native state^{28,29}. When unfolded rhodanese was diluted into a solution containing DnaK, DnaJ and Mg-ATP, addition of GroEL/ES after 10 min did not result in reactivation, however (Fig. 4*a*). We tested the possibility that GrpE was required to catalyse the transfer of rhodanese to GroEL. Indeed, upon addition of GrpE, there was rapid and efficient folding to the active enzyme. Although GrpE alone is unable to stabilize unfolded rhodanese in the aggregation assay (Fig. 1*b*), we do not exclude the possibility that an interaction between GrpE and the folding protein may be of functional importance in coupling the GroEL/ES-dependent folding reaction. When denatured rhodanese was diluted into buffer alone, the addition of any combination of heat-shock proteins and Mg-ATP after 10 min did not result in reactivation (not shown). When present on dilution, however, DnaK/DnaJ maintained rhodanese competent for folding by GroEL/ES for more than

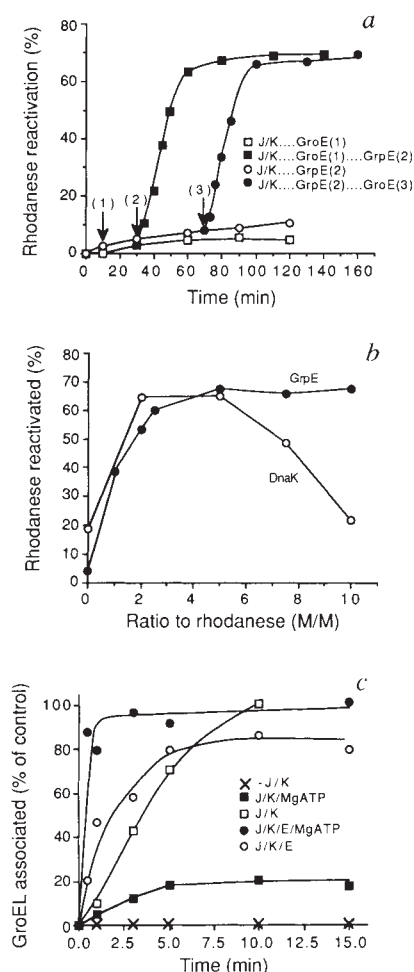


FIG. 4 GrpE-dependent transfer to GroEL and GroEL-mediated reactivation of rhodanese. **a**, Time course of reactivation of DnaK/DnaJ bound rhodanese in the presence of Mg-ATP after addition of GroEL/ES (GroE) (□) and GrpE (■) at 10 and 20 min following dilution of denatured rhodanese, respectively; reactivation in the presence DnaK, DnaJ, GrpE and Mg-ATP (○) and upon addition of GroEL/ES 70 min after dilution of rhodanese (●). The activity of native rhodanese is set to 100%. Activities measured upon spontaneous renaturation are subtracted. **b**, Effects of increasing concentrations of DnaK (○) and GrpE (●) in the presence of DnaJ on GroEL/ES-dependent reactivation. **c**, Association of rhodanese with GroEL after addition of GroEL to aggregated rhodanese (x) and to DnaK/DnaJ (K/J) associated rhodanese in the absence of GrpE (E) with (■) and without Mg-ATP (□), and in the presence of GrpE with (●) and without Mg-ATP (○). More than 95% of denatured rhodanese added directly to GroEL was recovered as GroEL bound (set to 100%).

METHODS. **a**, Rhodanese/DnaK/DnaJ complex (0.48 μ M/2.4 μ M/0.96 μ M) was generated as described in the legend to Fig. 1. ATP and magnesium were added to 1 mM and 5 mM, respectively. After preincubation at 25 °C for 10 min, GrpE (2.4 μ M) and GroEL/ES (0.48 μ M/0.96 μ M) were added when indicated. At various times aliquots were withdrawn and the activity of rhodanese was assayed in the presence of 10 mM *trans*-1,2-cyclohexanediamine-tetraacetate (CDTA) to stop ATP hydrolysis^{28,42}. Reactivation upon addition of GroEL/ES 10 min after diluting rhodanese into buffer lacking DnaK/DnaJ was equivalent to spontaneous reactivation (2–5% of the native enzyme control; not shown). **b**, Concentrations of DnaK and GrpE in the GroEL/ES dependent reactivation assay were varied at constant concentrations of GrpE (2.4 μ M), DnaK (2.4 μ M) and DnaJ (0.96 μ M), respectively. Activities were measured after 30 min of incubation with GroEL/ES. **c**, Rhodanese/DnaK/DnaJ complex (0.63 μ M/3.15 μ M/1.26 μ M) was incubated for 10 min at 25 °C. Then GroEL was added to 0.95 μ M. In one reaction, DnaK/DnaJ was omitted during preincubation (–J/K). When indicated, transfer reactions contained Mg-ATP and 3.15 μ M GrpE (E). At the times indicated, aliquots of the reactions received CDTA, were cooled on ice and were separated on 4–20% non-denaturing polyacrylamide gels^{21,52,63}. GroEL containing bands were excised and amounts of associated rhodanese were analysed by SDS-PAGE, immunoblotting and densitometry as in Fig. 3.

one hour even in the presence of GrpE and Mg-ATP (Fig. 4a).

The conditions for efficient transfer of rhodanese to GroEL and subsequent reactivation were analysed in more detail. Rhodanese bound to DnaJ alone could be reactivated by GroEL/ES with a yield of only 20% (Fig. 4b). GrpE did not affect this reaction, suggesting that the ATP hydrolytic activity of DnaK was necessary to drive the release of DnaJ. Reactivation was optimal at a 2–3-fold molar excess of DnaK over DnaJ (Fig. 4b), but at higher concentrations of DnaK the yield of active enzyme decreased. Transfer to GroEL also became less efficient when the concentration of DnaK exceeded that of GrpE. To separate the transfer reaction from GroEL-dependent folding, we used our earlier observation that rhodanese forms a complex with GroEL arrested in folding when GroES is absent²⁸. Co-electrophoresis of rhodanese and GroEL on non-denaturing polyacrylamide gels served as a direct measure of complex formation^{21,52}. On dilution from denaturant into a reaction containing a 1.5-fold molar excess of GroEL 14-mer, rhodanese was quantitatively recovered as the GroEL-bound species. No GroEL-bound protein was detected, however, when GroEL had been added to aggregated rhodanese (–J/K in Fig. 4c). Incubation of GroEL with the DnaK/DnaJ-associated protein and Mg-ATP resulted in the transfer of only 20% of total rhodanese. As expected, GrpE and Mg-ATP were necessary to allow rapid and fully efficient transfer to GroEL. The non-hydrolysable ATP analogue AMP-PNP did not support this reaction (not shown). When Mg-ATP was omitted rhodanese reached the chaperonin with slower kinetics, but independent of GrpE (Fig. 4c) because of the lower binding affinity of DnaK/DnaJ under these conditions (Fig. 1b). Interestingly, the GroEL/ES-mediated reactivation of this rhodanese also required GrpE. Apparently the Mg-ATP necessary for GroEL function increased the affinity of DnaK/DnaJ for the substrate, thus rendering substrate transfer and subsequent folding GrpE-dependent. It could be that complete release of rhodanese from DnaK/DnaJ takes place only upon interaction of the ternary protein complex with GroEL.

Catalytic cycles of folding

We tested whether maintaining the protein in an unfolded conformation by DnaK/DnaJ would allow GroEL to carry out multiple cycles of rhodanese folding. Previous studies had failed to demonstrate more than a single round of chaperonin action^{24–29}. In these experiments, the protein substrate had to be bound to stoichiometric amounts of GroEL 14-mer on dilution from denaturant. Rhodanese that did not stably interact with GroEL aggregated²⁸. The potential of GroEL/ES to act repeatedly could be demonstrated by adding saturating concentrations of denatured protein successively for two rounds of folding (Fig. 5a). When rhodanese was first stabilized by binding to DnaK/DnaJ, 4–5 rounds of reactivation could be observed per GroEL 14-mer in a process fully dependent on GrpE (Fig. 5b). Reactivation ceased when the transfer of rhodanese to GroEL became inefficient owing to the high excess of DnaK/DnaJ over GroEL in these experiments. The ability of GroEL to accept substrate molecules repeatedly from a pool of folding competent chains suggests that the cooperation between the GroEL/ES system and DnaK/DnaJ has the potential to function in a sequential folding pathway.

Discussion

We have described the reconstitution of a chaperone-mediated protein folding pathway from purified components. DnaK, DnaJ and GroEL can interact successively with a folding polypeptide, apparently directed by binding specificities for structural elements which are exposed sequentially during folding.

Studies *in vitro* have demonstrated that productive folding generally requires the presence of a complete polypeptide or folding domain⁸. *In vivo*, however, proteins emerge from ribosomes or from the *trans*-side of membranes as extended chains that are restricted in forming tertiary structural interactions and

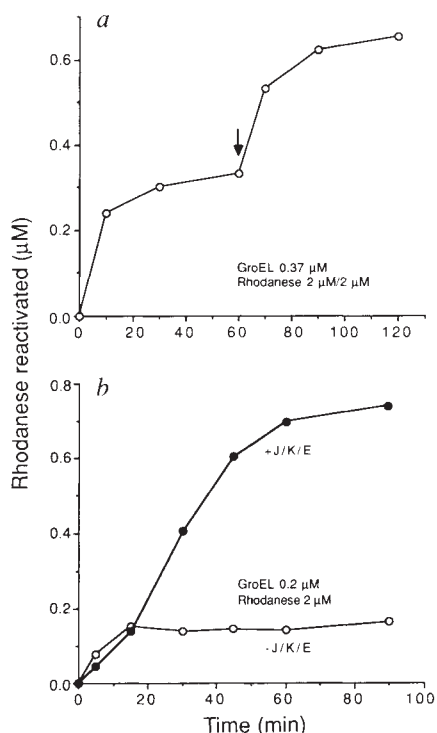


FIG. 5 Repeated cycles of GroEL action. *a*, Time course of reactivation after successive addition of rhodanese. Arrow indicates second addition of rhodanese. *b*, GroEL/ES-dependent reactivation of DnaK/DnaJ stabilized rhodanese in the presence of GrpE (●) or added directly to substoichiometric concentrations of GroEL (○).

METHODS. *a*, Guanidinium-HCl-denatured rhodanese was diluted to 2 μM into buffer B containing 0.37 μM GroEL. After 10 min at 25 °C, reactivation was started with 0.74 μM GroES and Mg-ATP. Activities were determined as described for Fig. 4. After 60 min (arrow), rhodanese was again added to 2 μM (from 200 μM stock in 6 M guanidinium-HCl). The concentration of reactivated rhodanese was calculated on the basis of the specific activity of the native enzyme. Spontaneous reactivation was negligible. *b*, Denatured rhodanese was diluted to 2 μM into buffer containing 10 μM DnaK (K) and 4 μM DnaJ (J). The complex was incubated for 10 min at 25 °C in the presence of GrpE (E) (10 μM) and Mg-ATP (+J/K/E). GroEL-dependent reactivation was started by the addition of GroEL/ES (0.2 μM/0.4 μM). The amount of enzyme reactivated by DnaK, DnaJ and GrpE alone (5–10% of total after 90 min) was determined in a reaction from which GroEL/ES was omitted. Subtraction of this background resulted in the curve shown representing GroEL/ES-dependent reactivation. The amount of renatured enzyme corresponding to a single GroEL-dependent folding cycle was determined by adding 2 μM denatured rhodanese directly to GroEL/ES containing buffer (–J/K/E).

lack stable secondary structure until a sufficient length of the polypeptide chain is reached. Incorrect folding and aggregation of incomplete (nascent) and complete newly made polypeptides may be prevented by members of the Hsp70 family^{10–13,30,35,36}, which have an affinity for synthetic peptides containing a threshold level of hydrophobic residues³¹. A complete sequence of chaperone reactions capable of accompanying a protein along the folding pathway was reproduced by using substrate proteins with different folding properties: R-CMLA, which maintains an extended conformation under folding conditions, and rhodanese, which rapidly collapses to a molten-globule-like state^{42,43} prone to aggregation⁴¹. DnaK recognized the extended R-CMLA, in contrast to DnaJ or GroEL, but failed to prevent efficiently the aggregation of rhodanese. The cooperation of DnaJ was required to stabilize this protein in an intermediate conformational state lacking the native tertiary structure.

Our results suggest that the following cascade of molecular chaperone activities might catalyse folding of nascent proteins (Fig. 6): First, DnaK interacts with extended sequences emerging from the ribosome (stage 1) (just as mitochondrial Hsp70 binds the polypeptides emerging from the inner mitochondrial membrane^{35,36}). Once an appropriate length of the polypeptide is synthesized and is made available for folding, the nascent chain will adopt the conformation of a collapsed folding intermediate (stage 2). This process might be promoted by slow ATP-dependent release of DnaK. Release of DnaK becomes limited by the formation of a DnaK/DnaJ/folding protein complex. This DnaK remains bound in the presence of Mg-ATP in a ternary complex (stage 3). We consider it possible that certain folding polypeptides may interact with DnaJ followed by DnaK. Transfer of the DnaK/DnaJ-bound protein to GroEL (stage 4) requires GrpE as the coupling factor. Finally, the protein acquires its native conformation mediated by GroEL in an ATP-dependent reaction (stage 5). GroEL strongly enhances the slow rate of ATP-dependent reactivation observed with DnaK, DnaJ and GrpE and is able to carry out multiple rounds of folding. In the functional pathway with the other chaperone proteins, GroEL may act as a cellular catalyst of protein folding.

Successive steps of this sequential chaperone reaction are highly coupled. DnaJ has a dual role: retarding the

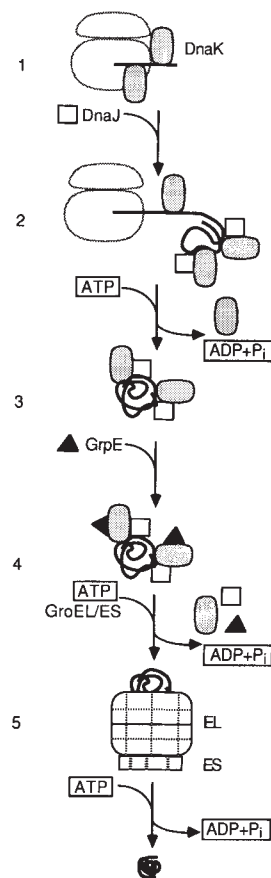


FIG. 6 Hypothetical model for the chaperone-mediated folding pathway of newly synthesized proteins (not drawn to scale). Binding of substrate protein and GroES to the same surface of GroEL is a possibility⁶⁸. The number of GroES bound per GroEL particle has not yet been determined. See 'discussion' for details.

ATP-dependent release of DnaK, and directly contributing to the stabilization of the folding protein. Consistent with this regulatory function, DnaJ is known to accelerate by about twofold the hydrolysis of DnaK-bound ATP, thus stabilizing the ADP-bound state of DnaK⁴⁶, which has high binding affinity for the unfolded protein³². GrpE drives the reaction forward by coupling the release of the substrate protein from DnaK/DnaJ to the next step in the pathway, the association with GroEL. The DnaK ATPase is stimulated ~50-fold (measured in the absence of bound protein) by DnaJ and GrpE acting synergistically, the latter facilitating ADP-ATP exchange⁴⁶. But productive release of a folding protein from DnaK/DnaJ requires the presence of GroEL. This may suggest that GroEL has to interact with the DnaK/DnaJ-associated substrate and/or with DnaK/DnaJ to trigger the transfer reaction. Alternatively, the folding protein may cycle between the chaperone-bound and free states as long as GroEL is missing. This could enable certain proteins to fold with the help of DnaK, DnaJ and GrpE alone, bypassing the GroEL/ES system.

For the folding pathway described here to be general for *E. coli*, all of the components should be represented by essential genes. Although this is the case for the known homologous chaperone proteins in mitochondria^{38,53}, the *dnaK* gene is only quasi-essential, in that a deletion grows slowly at 30°C and readily acquires suppressor mutations^{54,55}. Possibly GroEL/ES can function alone, albeit insufficiently. Alternatively, there might be a second DnaK/DnaJ-related pair of proteins able to function generally with low efficiency. As one might predict, *grpE* is essential at all growth temperatures⁵⁶. Lack of GrpE function may lead to the accumulation of folding proteins unable to dissociate from the ternary complex with DnaK and DnaJ.

DnaJ, DnaK and GrpE also cooperate in intracellular interac-

tions other than protein folding. DnaJ and DnaK were initially recognized by their role in bacteriophage λ replication¹⁶⁻¹⁸. Together with GrpE they function to disassemble the nucleoprotein structure required to localize the initiating DnaB helicase at the λ replication origin. In a related reaction, DnaJ associates with dimers of the RepA protein of phage P1, targeting DnaK to the complex; the resultant reaction is an ATP-dependent (but GrpE-independent) dissociation of RepA that allows for effective binding to origin DNA^{57,58}. These reactions appear to be distinct from the proposed functioning of DnaK and DnaJ in protein folding, in that DnaJ and DnaK interact with native proteins that require a conformational change for activation. Like the reaction in folding in both the λ and P1 reactions, however, the critical intermediate is probably a three-protein complex of DnaJ, DnaK, and λ P or P1 RepA proteins (refs 57, 58; and H. Hoffmann, S. Lyman and H.E., unpublished observations). The interaction of DnaJ with certain native target proteins, perhaps through exposed structural elements normally found in unfolded polypeptides, may result in partial unfolding and thus efficient recognition by DnaK. Similar mechanisms may be responsible for the renaturation of thermally denatured proteins¹⁵ which could re-enter a common folding pathway at the level of a DnaK/DnaJ-stabilized folding intermediate.

The principle of successive action of different chaperone components may be of more general importance for the folding of newly synthesized proteins and perhaps also for the refolding of damaged proteins. A functional cooperation of Hsp70s with DnaJ and GrpE-like proteins is likely to exist also in eukaryotes. This view is supported by the recent identification of various eukaryotic DnaJ homologues in different cellular locations^{39,59-61}. A candidate protein for an Hsp60-like function in the eukaryotic cytosol has been proposed⁶²⁻⁶⁴. □

Received 2 January; accepted 3 March 1992.

1. Anfinsen, C. B. *Science* **181**, 223-230 (1973).
2. Jaenicke, R. *Biochemistry* **30**, 3147-3161 (1991).
3. Ellis, R. J. *Nature* **328**, 378-379 (1987).
4. Rothman, J. E. *Cell* **59**, 591-601 (1989).
5. Pelham, H. R. B. *Nature* **332**, 776-777 (1988).
6. Gething, M. J. & Sambrook, J. *Nature* **355**, 33-45 (1992).
7. Hurlley, S. M. & Helenius, A. *Rev. Cell Biol.* **5**, 227-307 (1989).
8. Fischer, G. & Schmid, F. X. *Biochemistry* **29**, 2206-2212 (1990).
9. Hemmingsen, S. M. *et al.* *Nature* **333**, 330-334 (1988).
10. Deshaies, R. J., Koch, B. D., Werner-Washburne, M., Craig, E. A. & Schekman, R. *Nature* **332**, 800-805 (1988).
11. Chirico, W. J., Waters, M. G. & Blobel, G. *Nature* **332**, 805-810 (1988).
12. Zimmerman, R., Sagstetter, M., Lewis, M. J. & Pelham, H. R. B. *EMBO J.* **7**, 2875-2880 (1988).
13. Phillips, G. J. & Silhavy, T. J. *Nature* **344**, 882-884 (1990).
14. Skowryda, D., Georgopoulos, C. & Zyliz, M. *Cell* **62**, 939-944 (1990).
15. Gaitanaris, M. E., Papavassiliou, A. G., Rubock, P., Silverstein, S. J. & Gottesman, M. E. *Cell* **61**, 1013-1020 (1990).
16. Georgopoulos, C., Ang, D., Liberek, K. & Zyliz, M. in *Stress Proteins in Biology and Medicine* (eds Morimoto, T., Tissiers, A. & Georgopoulos, C.) 191-221 (Cold Spring Harbor Laboratory, New York, 1990).
17. Craig, E. A. & Gross, C. A. *Trends Biochem. Sci.* **16**, 135-139 (1991).
18. Echols, H. J. *biol. Chem.* **265**, 14697-14700 (1990).
19. Cheng, M. Y. *et al.* *Nature* **337**, 620-625 (1989).
20. Goloubinoff, P., Gatenby, A. A. & Lorimer, G. H. *Nature* **337**, 44-47 (1989).
21. Ostermann, J., Horwich, A. L., Neupert, W. & Hartl, F.-U. *Nature* **341**, 125-130 (1989).
22. Barraclough, R. & Ellis, R. J. *Biochim. biophys. Acta* **608**, 19-31 (1980).
23. Bochkareva, E. S., Lissin, N. M. & Girshovich, A. S. *Nature* **336**, 254-257 (1988).
24. Goloubinoff, P., Christeller, J. T., Gatenby, A. A. & Lorimer, G. H. *Nature* **342**, 884-889 (1991).
25. Laminet, A. A., Ziegelhoffer, T., Georgopoulos, C. & Plückthun, A. *EMBO J.* **9**, 2315-2319 (1990).
26. Buchner, J. *et al.* *Biochemistry* **30**, 1586-1591 (1991).
27. Badcoe, E. *Biochemistry* **30**, 9195-9200 (1991).
28. Martin, J. *et al.* *Nature* **352**, 36-42 (1991).
29. Mendoza, J. A., Rogers, E., Lorimer, G. H. & Horowitz, P. M. *J. biol. Chem.* **266**, 13044-13049 (1991).
30. Beckman, R. P., Mizzen, L. A. & Welch, W. J. *Science* **248**, 850-854 (1990).
31. Flynn, G. C., Pohl, J., Flocco, M. T. & Rothman, J. E. *Nature* **353**, 726-730 (1991).
32. Palleros, D. R., Welch, W. J. & Fink, A. L. *Proc. natn. Acad. Sci. U.S.A.* **88**, 5719-5723 (1991).
33. Landry, S. J. & Gierasch, L. M. *Biochemistry* **30**, 7359-7362 (1991).
34. Rassow, J., Hartl, F. U., Guiraud, B., Pfanner, N. & Neupert, W. *FEBS Lett.* **275**, 190-194 (1990).
35. Kang, P.-J. *et al.* *Nature* **348**, 137-143 (1990).
36. Scherer, P. E., Krieg, U. C., Hwang, S. T., Vestweber, D. & Schatz, G. *EMBO J.* **9**, 4315-4322 (1990).
37. Manning-Krieg, U. C., Scherer, P. E. & Schatz, G. *EMBO J.* **10**, 3273-3280 (1991).
38. Reading, D. S., Hallberg, R. L. & Myers, A. M. *Nature* **337**, 655-659 (1989).
39. Blumberg, R. P. & Silver, P. *Nature* **349**, 627-630 (1991).
40. Ploegman, J. H. *et al.* *Nature* **273**, 124-129 (1978).
41. Tandon, S. & Horowitz, P. M. *J. biol. Chem.* **261**, 15615-15681 (1986).
42. Tandon, S. & Horowitz, P. M. *J. biol. Chem.* **264**, 9859-9866 (1989).
43. Tandon, S. & Horowitz, P. M. *J. biol. Chem.* **265**, 5967-5970 (1990).
44. Waxman, L. & Goldberg, A. L. *Science* **232**, 500-503 (1986).
45. Ostoa-Saloma, P., Ramirez, J. & Perez-Monfort, R. *Biochem. biophys. Acta* **1041**, 140-152 (1990).
46. Liberek, K., Marszałek, J., Ang, D., Georgopoulos, C. & Zyliz, M. *Proc. natn. Acad. Sci. U.S.A.* **88**, 2874-2878 (1991).
47. Sheffield, W. P., Shore, G. C. & Randall, S. K. *J. biol. Chem.* **265**, 11069-11076 (1990).
48. Christenson, H. & Pain, R. H. *Eur. Biophys. J.* **19**, 221-229 (1991).
49. Pittsyn, O. B., Pain, R., Semisotnov, G. V., Zernovnik, E. & Razgulyaev, U. I. *FEBS Lett.* **262**, 20-24 (1990).
50. Ewbank, J. J. & Creighton, T. E. *Nature* **350**, 518-520 (1991).
51. Bardwell, J. C. A., Tilly, K., Craig, E., King, J., Zyliz, M. & Georgopoulos, C. *J. biol. Chem.* **261**, 1782-1785 (1986).
52. Musgrove, J. E., Johnson, R. A. & Ellis, R. J. *Eur. J. Biochem.* **163**, 529-534 (1987).
53. Craig, E. A., Kramer, J. & Kosic-Smithers, J. *Proc. natn. Acad. Sci. U.S.A.* **84**, 4156-4160 (1987).
54. Sell, S. M., Eisen, C. Ang, D., Zyliz, M. & Georgopoulos, C. *J. bact.* **172**, 4827-4835 (1990).
55. Bukau, B. & Walker, G. C. *EMBO J.* **9**, 4027-4036 (1990).
56. Ang, D. & Georgopoulos, C. *J. Bact.* **171**, 2748-2755 (1989).
57. Wickner, S., Hoskins, J. & McKenney, K. *Nature* **350**, 165-167 (1991).
58. Wickner, S. *Proc. natn. Acad. Sci. U.S.A.* **88**, 7903-7907 (1991).
59. Sadler, I., Chiang, A., Kurihara, T., Rothblatt, J., Way, J. & Silver, P. *J. Cell Biol.* **109**, 2665-2775 (1989).
60. Caplan, A. J. & Douglas, M. G. *J. Cell Biol.* **114**, 609-621 (1991).
61. Luke, M. M., Sutton, A. & Arndt, K. T. *J. Cell Biol.* **114**, 623-638 (1991).
62. Ellis, R. J. *Science* **250**, 954-959 (1990).
63. Trent, J. D., Nimmegern, E., Wall, J. S., Hartl, F.-U. & Horwich, A. L. *Nature* **354**, 490-493 (1991).
64. North, G. *Nature* **354**, 434-435 (1991).
65. Welch, W. J. & Feramisco, J. R. *Molec. cell. Biol.* **5**, 1229-1237 (1985).
66. Bradford, M. M. *Analyt. Biochem.* **72**, 248-254 (1976).
67. Jaenicke, R. & Rudolph, R. in *Protein Structure, a Practical Approach* (ed. Creighton, T. E.) 191-222 IRL, Oxford, (1989).
68. Creighton, T. E. *Nature* **352**, 17-18 (1991).

ACKNOWLEDGEMENTS. We thank C. Georgopoulos for *E. coli* strains overproducing DnaJ, GrpE, and GroEL/ES, J. Ewbank and T. Creighton for purified R-CMLA, and H. Gemeinhardt for technical assistance. A large part of this research was done at the Institute of Physiological Chemistry, University of Munich. This work was supported by the Deutsche Forschungsgemeinschaft and by grants from the NIH.

Identification of the pulsar PSR1509–58 with the ‘guest star’ of AD 185

S. E. Thorsett

Owens Valley Radio Astronomy Observatory, California Institute of Technology 105–24, Pasadena, California 91125, USA

ON 7 December 185, astronomers at the imperial observatory of Lo-Yang, in central China, reported a ‘guest star’ in the southern sky. Their records of its appearance, gradual fading and disappearance constitute the oldest compelling historical account of a supernova. It is widely believed that the remnant of this explosion is MSH14–63, but the identification is problematic, in particular because the position of MSH14–63 with respect to the Sun and the horizon would have made it invisible for some of the time when the guest star was said to be visible. I argue here that the Chinese astronomers actually witnessed the birth of the pulsar PSR1509–58 in the supernova remnant MSH15–52. This is then only the second pulsar, after that in the Crab nebula (PSR0531+21), to have a known age. Timing measurements of the age of PSR1509–58 are consistent with its birth $\sim 1,800$ years ago, and future measurements of the distance to MSH15–52 and its rate of expansion should conclusively establish or disprove this identification.

Because of the great rarity of nearby supernovae, records kept by ancient astronomers form a uniquely valuable resource for modern scientists. Before SN1987A, there are good records of only seven events, in 185, 393, 1006, 1054, 1181, 1572 and 1604, and less-convincing evidence for a few more. Of these, only the supernovae of 1006, 1572 and 1604 were observed by European astronomers; by far the most valuable records come from the Far East. For an excellent historical review, see ref. 1.

The five historical supernovae of the past millennium have all been convincingly associated with known supernova remnants. The most famous such association is, of course, that between the guest star of 1054 and the Crab nebula, with its pulsar PSR0531+21 (refs 2–4). Plausible associations have been suggested for the earlier two events¹, although we will argue below that for the AD 185 event the case is far from settled.

As a much older event, it is not surprising that relatively little documentation concerning SN185 has survived. There is only a single reference in the *Hou-han-shu*, the official history of the Later Han dynasty (AD 23–220)⁵:

In the second year of the epoch Chung Ping, the 10th moon, on the day Kwei Hae [7 December 185], a strange star appeared in the middle of Nan-Mên [‘Southern Gate’]. It was like a large bamboo mat. It displayed the five colors, both pleasing and otherwise. It gradually lessened. In the 6th moon of the succeeding year [5 July–2 August 186] it disappeared.

On several points, the translation is ambiguous. The word *hounien*, here given its ancient meaning of ‘succeeding year’⁶, might also take its modern meaning, ‘the year after next’, in which case the star disappeared between 24 June and 23 July 187 (ref. 1). The asterism Nan-Mên, no longer visible from China, was the southernmost constellation visible from Lo-Yang (34.7° N) in AD 185. It is widely believed to have consisted of the two bright stars α and β Cen¹, although a few authors suggest ϵ and β Cen⁶. The latter belief seems to be founded largely on the star maps of Ho⁷, which extend only to a declination of -60° (epoch 1950) and which therefore erroneously exclude the very bright α Cen. Finally, the word *chung*, translated here as ‘middle’, may also indicate the more general ‘within’¹. It is possible that a more detailed description of the star’s position relative to the asterism has been lost, as the historiographers at the end of each dynasty often only summarized the astronomers’ daily reports.

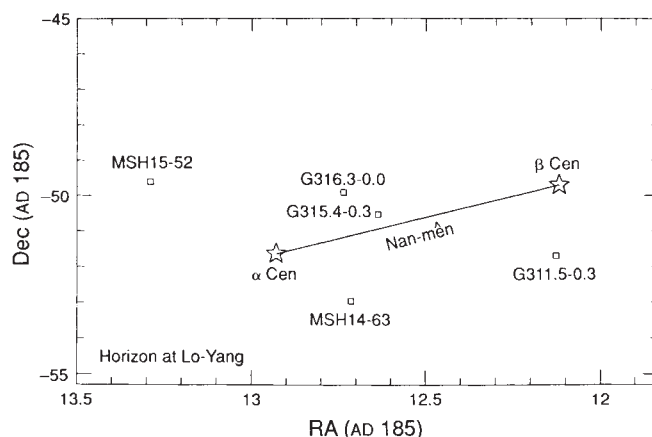


FIG. 1 Five candidate supernova remnants for SN185 (ref. 1). Positions are shown at epoch AD 185. Proper motion has been included for the stars but is expected to be negligible for the remnants.

Despite these limitations, the record has sufficient detail that it is sensible to ask which supernova remnant can be associated with the historical event. Clark and Stephenson limit their attention to the region between α and β Cen, and propose four candidate remnants¹ (Fig. 1). Three of these, G311.5–0.3, G315.4–0.3 and G316.3–0.0, can be excluded from consideration because of their great distances¹. Clark and Stephenson suggest the remaining source, MSH14–63, as the probable remnant of SN185.

This identification is problematic for several reasons, most important of which is that the fading supernova would not have been visible during the sixth month of AD 186 when it is reported to have vanished¹. On the first day of that month, 5 July, the star would already have disappeared into the twilight, with an elevation at sunset of less than 0.8° . (This problem persists if the year of disappearance is taken to be AD 187). Clark and Stephenson, recognizing the difficulty, suggested an error of one month in the historical records, but given the extreme precision of most dates, this seems unlikely¹.

A second problem is that the remnant seems to be too distant. Unfortunately, no H I absorption measurement seems to exist, but an OB association⁸, an optical absorption measurement⁹ and the (less reliable¹⁰) $\Sigma - D$ measurement¹¹ agree on a distance of 2.5–3.2 kpc. Several arguments based on the ‘known’ age of the remnant, however, require it to be much closer. For example, measurements of Balmer-dominated optical filaments¹² show a shock speed of $\sim 800 \text{ km s}^{-1}$. Using the known angular diameter, $\sim 40'$, and assuming Sedov phase evolution, it can be shown that $D_{\text{kpc}} \approx 0.3 \tau_{\text{kyr}}$, where D_{kpc} is the remnant distance in kiloparsecs and τ_{kyr} is the age of the remnant in thousands of years. Assuming a distance of 3 kpc, the remnant age is ~ 10 kyr, much too old to account for the SN185 event. (The inferred age would be ~ 4 kyr if the remnant is still in free expansion.) Results from modelling of the X-ray emission are similar. If the age of the remnant is constrained to be 1.8 kyr, then the distance must be only ~ 1 kpc (ref. 13), whereas if the distance is set to the measured value of 3 kpc, then the age of the remnant is about 6 kyr.

Faced with a paucity of plausible candidates between the two stars of Nan-mên, we turn instead to one just outside: MSH15–52, located just 3° from α Cen (Fig. 1). MSH15–52, also known as G320.4–1.2, is a partial shell with diameter $\sim 30'$. The north-west part of the shell contains a bright optical knot, RCW86, which is also a soft X-ray source. A summary of its radio, optical and X-ray properties can be found in ref. 14. A H I absorption measurement¹⁵ suggests a distance of ~ 4.2 kpc. At the centre of the remnant is a pulsar, visible at both X-ray¹⁶ and radio¹⁷ wavelengths, embedded in a synchrotron nebula¹⁸.

The timing properties of the pulsar, PSR1509–58, greatly strengthen the identification of MSH15–52 with the supernova of AD 185. It is the second-youngest known radio pulsar after the Crab pulsar. Given the measured pulsar frequency ν , frequency derivative $\dot{\nu}$ and braking index n , we can calculate the characteristic age (in 1982) of the pulsar, $\tau = -\nu/(n-1)\dot{\nu} = 1.69$ kyr (ref. 19). The true age may be slightly greater if the braking index increases with time, or less if the initial spin frequency was not much greater than the current frequency. Without an assumption of a variable braking index, the characteristic age of the pulsar excludes the 1.8-kyr-old supernova at about the 2σ level¹⁹, but this is almost certainly due to a slight corruption of the braking index by red timing noise such as that observed in nearly all young pulsars²⁰. In any case, it is hard to imagine that the close match between the pulsar age and that of SN185 can be coincidental. Even with the most generous estimates of the galactic supernova rate, the chances of two different supernovae in the direction of α and β Cen within a few hundred years are remote.

Despite the evident youth of PSR1509–58, its remnant is sometimes assigned an age of ~ 10 kyr, on the basis of a model of Sedov expansion into a moderately dense (1 cm^{-3}) medium¹⁴. Measurements of the optical filaments in the bright knot RCW86 also indicate that neither free nor adiabatic expansion is consistent with their slow velocities²¹. There is no problem reconciling the remnant and pulsar ages, however, if the remnant's initial expansion was into an underdense medium ($\sim 0.01\text{ cm}^{-3}$) such as might be expected in the stellar wind bubble of the massive progenitor star^{14,22,23}. Interaction with a high-density medium on the northwest side (towards the galactic plane) would account for both the enhanced luminosity and the slow expansion of the knots.

It is straightforward to make a rough estimate of how bright the supernova would have been in China. If we assume an absolute magnitude at supernova peak of about -19 , a distance of 4.2 kpc and an absorption $A_V \approx 3$ (absorption varies across the remnant; this is a typical value²⁴), the apparent magnitude at peak would be about -3 , similar to a bright planet. With atmospheric extinction, due to its low elevation, the apparent magnitude from Lo-Yang would be about -1 .

On 7 December 185, MSH15–52 was only about a week past its heliacal rising, and was at an elevation of $\sim 2.9^\circ$ at sunrise, high enough for an object of magnitude $m = -1$ to be observed as a morning star. Note that there is no indication in the record that SN185 was ever observable during daylight hours. The coincidence of the reported date of appearance of the guest star and its heliacal rise date suggests that the supernova probably occurred during the weeks (or months) before 7 December, but was unobservable until then. Low elevation (and possibly weather) may account for the week's delay of its discovery after heliacal rising.

On 5 July 186, MSH15–52 was still at an elevation of $\sim 4.2^\circ$ at sunset, and was above the horizon for 1.5 hours after sunset, so it would have been easy to observe as an evening star. By the end of the month, however, the source set before the Sun, hence its disappearance. No error in the record need be assumed. By the second heliacal rise date, a year after the initial discovery, the guest star would have faded below naked-eye visibility.

Although the good agreement of the heliacal rising and setting dates of MSH15–52, together with the age of the pulsar PSR1509–58, form a compelling case for the identification of MSH15–52 with SN185, this identification is not conclusive. Several observations are desired. The first are high-resolution radio maps of both MSH15–52 and MSH14–63 at several epochs, to measure the proper expansion of both remnants. In both cases, an age of 1,800 years suggests an expansion rate of ~ 1 arcsec per year, or a few times smaller if the remnant has recently gone from free to adiabatic expansion. It is important to measure the expansion of the low-luminosity radio shells, rather than the higher-luminosity filaments which may arise in

high-density clouds. Second is a measurement of the distance to MSH14–63, by H I absorption if possible. A firm distance limit of more than ~ 2 kpc will rule it out as a candidate.

The identification of MSH15–52 with SN185 is interesting history, but also important astronomy. MSH15–52 is a thousand years older than the next-oldest supernova for which a secure age exists; its study will provide insight into the evolution of remnants in early middle-age. Published studies of PSR1509–58 include less than 2 years of timing data¹⁹; with a decade now since its discovery it should be possible to refine the $\dot{\nu}$ measurement of PSR1509–58, and hence its 'timing age'. This will have implications for the early time evolution of pulsar braking indices, or, possibly, the initial spin period of the neutron star. $\square$

Received 3 February; accepted 9 March 1992.

1. Clark, D. H. & Stephenson, F. R. *The Historical Supernovae* (Pergamon, Oxford, 1977).
2. Lundmark, K. *Publ. astr. Soc. Pacific* **33**, 225–238 (1921).
3. Duyvendak, J. J. L. *Publ. astr. Soc. Pacific* **54**, 91–94 (1942).
4. Mayall, N. U. & Oort, J. H. *Publ. astr. Soc. Pacific* **54**, 95–104 (1942).
5. Williams, J. *Observations of Comets: From 611 B.C. to A.D. 1640* (London, 1871).
6. Huang, Y.-L. & Moriarty-Schieven, G. H. *Science* **235**, 59–60 (1987).
7. Ho, P. Y. *The Astronomical Chapters of the Chin Shu* (Mouton, Paris, 1966).
8. Westerlund, B. E. *Astr. J.* **74**, 879 (1969).
9. Leibowitz, E. M. & Danziger, I. J. *Mon. Not. R. astr. Soc.* **204**, 273 (1983).
10. Green, D. A. *Mon. Not. R. astr. Soc.* **209**, 449–478 (1984).
11. Clark, D. H. & Caswell, J. L. *Mon. Not. R. astr. Soc.* **174**, 267–305 (1976).
12. Lyne, A. G. & Bailes, M. *Mon. Not. R. astr. Soc.* **246**, 15–17 (1990).
13. Claas, J. J., Smith, A., Kaastra, J. S., de Korte, P. A. J. & Peacock, A. *Astrophys. J.* **337**, 399–407 (1989).
14. Seward, F. D., Harnden, F. R., Murdin, P. & Clark, D. H. *Astrophys. J.* **267**, 698–710 (1983).
15. Caswell, J. L., Murray, J. D., Roger, R. S., Cole, D. J. & Cooke, D. J. *Astr. Astrophys.* **45**, 239 (1975).
16. Seward, F. D. & Harnden, F. R. *Astrophys. J.* **256**, L45–L47 (1982).
17. Manchester, R. N., Tuohy, I. R. & D'Amico, N. *Astrophys. J.* **262**, L31–L33 (1982).
18. Seward, F. D., Harnden, F. R., Szymkowiak, A. & Swank, J. *Astrophys. J.* **281**, 650–657 (1984).
19. Manchester, R. N., Durbin, J. M. & Newton, L. M. *Nature* **313**, 374–376 (1985).
20. Manchester, R. N. & Taylor, J. H. *Pulsars* (Freeman, San Francisco, 1977).
21. van den Bergh, S. & Kamper, K. W. *Astrophys. J.* **280**, L51–L54 (1984).
22. Srinivasan, G., Dwarakanath, K. S. & Radhakrishnan, V. *Curr. Sci.* **51**, 596 (1982).
23. Bhattacharya, D. *J. Astrophys. Astr.* **11**, 125–140 (1990).
24. Lortet, M. C., Georgelin, Y. P. & Georgelin, Y. M. *Astr. Astrophys.* **180**, 65–72 (1987).

ACKNOWLEDGEMENTS. I thank S. Kulkarni for helpful conversations, and V. Kaspi, H. Johnston and J. A. Phillips for comments on the manuscript. This work was supported by a Robert A. Millikan Research Fellowship.

Numerical calculations of fast dynamos in smooth velocity fields with realistic diffusion

D. J. Galloway* & M. R. E. Proctor

Department of Applied Mathematics and Theoretical Physics,
University of Cambridge, Silver Street, Cambridge CB3 9EW, UK

MANY astrophysical magnetic fields are thought to arise by dynamo action due to internal fluid motions, but the natural timescale for magnetic field growth is the diffusion timescale, which in realistic astrophysical applications is very large¹. A fast dynamo is one that operates on the much shorter turnover timescale of the generating fluid flow, and the analytical intractability of smooth flows with diffusion has prompted the use of many ingenious models^{2–10}, differing from the true problem in having a modified or time-dependent diffusion or singularities in the flow field. Here we adopt a straightforward approach and present numerical computations of linear kinematic dynamos associated with periodic smooth flows, with diffusion explicitly included. Examples of time-varying flows depending on two spatial coordinates give convincing evidence of fast dynamo action for diffusion times up to 10,000 times greater than the turnover time. A three-dimensional steady flow shows similar behaviour, although computations have not been

* Permanent address: Research Centre for Theoretical Astrophysics and School of Mathematics and Statistics, University of Sydney, New South Wales 2006, Australia.

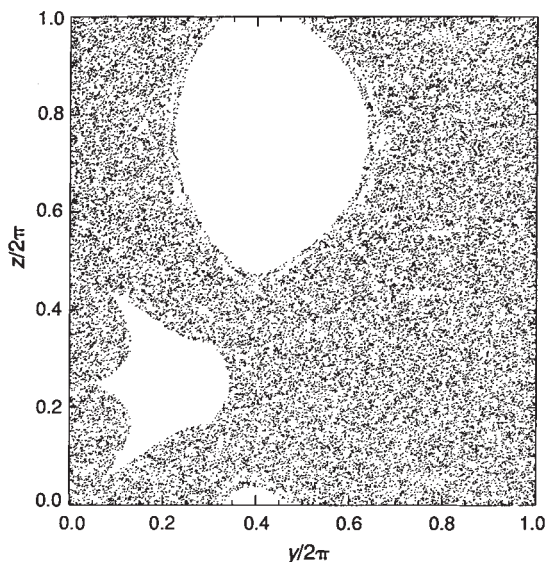


FIG. 1 Poincaré section for the CP flow at $t=0$ in the $(y-z)$ plane (the x -component of $\mathbf{u}$ is suppressed). Points are plotted at equal time intervals $2\pi/\omega$, and y and z values are plotted modulo 2π .

carried out so far and the asymptotic behaviour is less clear. All these flows have large regions where particle paths are chaotic.

As a simple model of the dynamo process, we prescribe a velocity field $\mathbf{u}$ and seek exponentially growing solutions to the induction equation for the magnetic field $\mathbf{B}$

$$\frac{\partial \mathbf{B}}{\partial t} = \nabla \times (\mathbf{u} \times \mathbf{B}) + \frac{1}{R_m} \nabla^2 \mathbf{B}$$

This is known as the kinematic dynamo problem^{11,12}. The equation has been scaled with a characteristic length L , velocity U and time L/U . The quantity $R_m = UL\mu_0\sigma$ is a nondimensional measure of the fluid conductivity σ , and is enormous in astrophysics (typically of order 10^{10}). With this scaling, the flow turnover time is of order 1 and the electromagnetic diffusion time of order R_m .

Although many flows are known that give a growing $\mathbf{B}$ (refs 11, 12), nearly all of these have the feature that the growth rate tends to zero as $R_m \rightarrow \infty$; such flows have been termed slow dynamos¹. These are supposedly ineffective in astrophysics: the resulting fields would have grown insufficiently during the age of the Universe¹, and short-term variations such as the solar cycle are hard to explain with slow dynamos. Accordingly, attention has now focused on the possibility of fast dynamos, where the growth rate becomes independent of R_m as $R_m \rightarrow \infty$. Because analysis has proved exceptionally difficult in this (singular) limit, various adjustments have been made to the problem to help in detecting the asymptotic behaviour. Singularities have been allowed in the velocity⁴ and the vorticity⁵; in these cases the speed of the resulting dynamos is apparently due to the singularities themselves. Velocity fields that are discontinuous in time² and intermittent diffusion rates^{2,3} have also been used; these are less controversial, although a formal proof that the solutions merge with those of the induction equation in the large R_m limit remains elusive. Other work has focused on solutions of the diffusionless equations^{7,8}; again, there is no formal connection with the finite-diffusivity case. A more recent approach⁹ uses a stochastic diffusion model, integrating along the particle paths (for a review, see ref. 10). A numerical demonstration of fast dynamo action for two pulsed Beltrami vortices was reported by Otani¹⁴, but no details have subsequently been published. Thus there still seems to be scope for high-accuracy numerical experiments with judiciously chosen velocity fields.

Motions that are chaotic are prime candidates for fast dynamos. This is because in the perfectly conducting case, magnetic field lines are advected like fluid line elements^{11,13}. Chaotic flows stretch these exponentially, providing a powerful mechanism for growth of the field. Indeed, flows with no chaos and no hyperbolic stagnation points always yield slow

dynamos²⁰. The 'ABC' flows¹⁵

$$\mathbf{u} = (A \sin z + C \cos y, B \sin x + A \cos z, C \sin y + B \cos x)$$

where A , B and C are constants, are known to possess chaotic regions when $ABC \neq 0$ (refs 16, 17), and their dynamos have been examined to see whether they are fast, with inconclusive results^{18,19}. Research has mainly concentrated on the case $A = B = C$, although unfortunately this special case has very small chaotic regions.

Chaotic flows and their induced fields are by their nature highly resistant to analytical methods. On the other hand, numerical methods cannot currently resolve values of R_m higher than a few hundred in three space dimensions¹⁹ and this is not really high enough. It is known, however, that flows depending on only two space dimensions can be chaotic if they are time-dependent. Their dynamos can be investigated with a two-dimensional code, thus enabling much higher values of R_m to be attained. Accordingly, let us consider modifying the integrable ABC flow with $B=0$ so that its phasing depends on time in two possible ways:

$$\mathbf{u} = (A \sin(z + \sin \omega t) + C \cos(y + \cos \omega t), \\ A \cos(z + \sin \omega t), C \sin(y + \cos \omega t))$$

$$\mathbf{u} = (A \sin(z + \cos \omega t) + C \cos(y + \cos \omega t), \\ A \cos(z + \cos \omega t), C \sin(y + \cos \omega t))$$

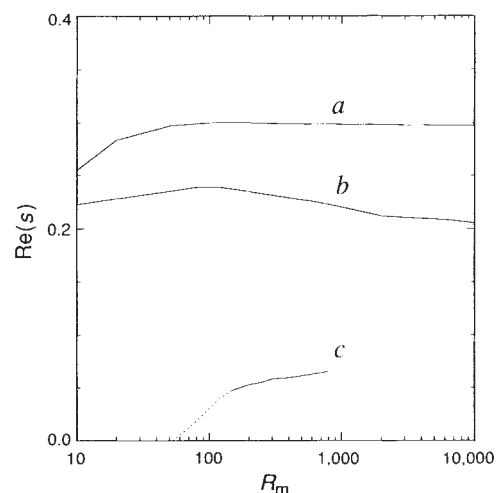


FIG. 2 Real part of the growth rate S plotted against R_m (log scale) for the three flows discussed in the text. *a*, CP flow, $k=0.57$. *b*, LP flow, $k=0.62$. *c*, three-dimensional flow. The dotted portion of the latter is determined with lower accuracy.

These are now non-integrable and possess large chaotic regions (see Fig. 1). The first is 'circularly polarized' (CP) and stirs the basic $A(B)C$ flow around, whereas the second is 'linearly polarized' (LP), and shakes it back and forth parallel to the $y = z$ diagonal. In what follows we treat only the case $A = C = \sqrt{3}/2$, $\omega = 1$.

The induction equation is solved numerically by writing the magnetic field as a triple Fourier series and timestepping the Fourier amplitudes; the method is a straightforward adaptation of that in ref. 19. Because $B = 0$, each mode proportional to e^{ikx} evolves independently of modes with different k . Thus we can fix k and compute the resulting $\mathbf{B}$ with a two-dimensional code. The y - z periodicity of the solution is taken for convenience to be the same as that of the flow. As k is a free parameter, one has to optimize over k to find the value k_c giving the eigenvalue with largest real part. The solution with zero k is known to decay from the two-dimensional analogue of Cowling's theorem¹; for large enough k we should find decay, although k_c may grow with R_m . This is seen in Soward's⁵ infinite-vorticity fast dynamo, which uses a very similar but steady velocity field, and for which $k_c \approx R_m^{1/2}$. Surprisingly, k_c is here of order 1 and varies little if at all with R_m , being 0.57 (CP flow) and 0.62 (LP flow): the behaviour of the real part of the growth rate as a function of R_m for the two flows is shown in Fig. 2. The CP flow in particular provides convincing evidence for fast dynamo action, the real part scarcely changing between $R_m = 50$ and $R_m = 10,000$. The LP flow is only marginally less convincing. The real parts of the growth rates for the magnetic energy tend to 0.3 and 0.2 respectively; there are only very small fluctuations in the energy due to the temporal variation of the velocity field. For comparison, the largest Lyapunov exponents in the chaotic regions are 1.43 and 1.17. All Fourier amplitudes are initially filled equally, and as time t advances, there is a period where the energy falls as the high-order modes decay. For large enough R_m , the energy levels off at $t \approx 10$ and exponential growth begins at $t \approx 25$. Thus the form of the eigenfunction seems to be established on the turnover timescale. This was confirmed by running the case $R_m = 400$ for a whole diffusion time. The code itself was checked against a multiple-timescale analysis at low R_m .

The structure of the eigenfunction for the CP flow is shown in Fig. 3 (the LP solution looks rather similar but without the structures parallel to $y = z$). A detailed comparison with the velocity field is difficult because of the three-dimensionality, but broadly speaking the field structures line up along the unstable directions of stagnation points of the y - z flow. The thickness of the structures goes down as R_m increases, presumably scaling²¹ as $R_m^{-1/2}$, the behaviour found in analogous, analytically tractable situations^{22,23} (although here the structure may be fractally plated in the limit of infinite R_m). As k is of order 1, the structures are like curved sheets. As $R_m \rightarrow \infty$ these eigenfunctions must tend to generalized functions although, remarkably, their growth rate tends to a finite limit.

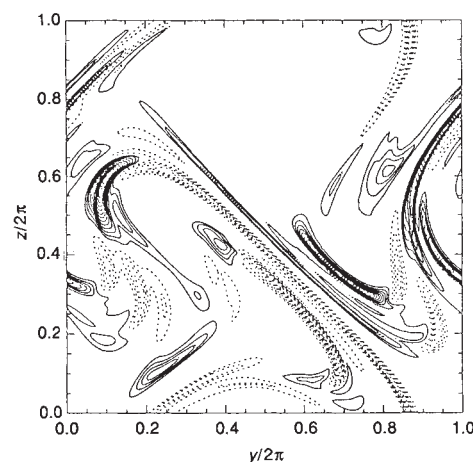


FIG. 3 Contours of B_x at $x=0$: CP flow, $k=0.57$, $R_m=2,000$.

Finally we give results for the three-dimensional steady flow $\mathbf{u} = (\sin z, \sin x, \sin y)$, which is the $A = B = C = 1$ flow with the cosines omitted, making it roughly twice as efficient to compute (but still requiring a three-dimensional code). Poincaré sections reveal that almost all the flow region is chaotic, but in contrast to the ABC flow, the helicity density $\mathbf{u} \cdot \nabla \times \mathbf{u}$ has zero spatial average. This is supposedly un conducive to dynamo action but, as shown in Fig. 2, the flow operates perfectly satisfactorily as a dynamo, and in fact seems likely to be fast. The imaginary part of the growth rate is zero. As yet we have only achieved results for R_m up to 800 (using $(96)^3$ resolution), with no information on eigenfunction structure. Further computations are planned, and we should be able to obtain resolved solutions for values of R_m up to 2,000.

These computations provide strong evidence for the existence of fast dynamos, but no proof. For the astrophysicist such a proof is probably academic: all his flows are likely to be highly turbulent and chaotic, and the dynamos that they produce seem likely to be fast. Calculating the fields directly seems possible only by numerical simulation or by the use of theories that parameterize fast dynamo generation (for example, the α -effect¹). A more practical challenge is to translate fast dynamo ideas and models to more realistic geometries such as spheres and discs surrounded by empty space. Another is to examine the back reaction of the field on the motion. This is interesting because there is a physical inconsistency inherent in kinematic fast dynamos: elementary consideration of the effects of field amplification suggests that however small the seed field, for large R_m the Lorentz force becomes significant after a time much shorter than the diffusion time, albeit over very small volumes. In fact, intermittent fields seem to be common in astrophysics, with strong fields concentrated in small regions; a sunspot is an obvious example. $\square$

Received 3 December 1991; accepted 13 March 1992.

1. Zel'dovich, Ya. B., Ruzmaikin, A. A. & Sokoloff, D. D. *Magnetic Fields in Astrophysics* (Gordon and Breach, New York, 1983).
2. Bayly, B. J. & Childress, S. *Geophys. Astrophys. Fluid Dyn.* **44**, 211–240 (1988).
3. Finn, J. M., Hanson, J. D., Kan, I. & Ott, E. *Phys. Rev. Lett.* **62**, 2965–2968 (1989).
4. Gilbert, A. D. *Geophys. Astrophys. Fluid Dyn.* **44**, 214–258 (1988).
5. Soward, A. M. *J. Fluid Mech.* **180**, 267–295 (1987).
6. Finn, J. M., Hanson, J. D., Kan, I. & Ott, E. *Physics Fluids* **B3**, 1250–1269 (1991).
7. Gilbert, A. D. *Nature* **350**, 483–485 (1991).
8. Gilbert, A. D. *Phil. Trans. R. Soc.* (submitted).
9. Klapper, I. *J. Fluid Mech.* (in the press).
10. Childress, S. *Fast Dynamo Theory*, Proc. Workshop on Topological Fluid Dynamics, Santa Barbara (eds Moffatt, H. K. & Tabor, M.) (in the press).
11. Moffatt, H. K. *Magnetic Field Generation in Electrically conducting Fluids* (Cambridge University Press, 1978).

12. Roberts, P. H. & Soward, A. M. *Rev. Fluid Mech.* **24**, 459–512 (1992).
13. Arnold, V. I., Zel'dovich, Ya. B., Ruzmaikin, A. A. & Sokoloff, D. D. *Zh. éksp. teor. Fiz.* **81**, 2052–2058 (1981); (Engl. transl.) *Sov. Phys. JETP* **54**, 1083–1086 (1981).
14. Otani, N. *Eos* (abstr.) **69**, 1366 (1989).
15. Arnold, V. I. *C.R. hebd. Séanc. Acad. Sci., Paris* **261**, 17–20 (1965).
16. Hénon, M. *C.R. hebd. Séanc. Acad. Sci., Paris* **262**, 312–314 (1966).
17. Dombre, T. et al. *J. Fluid Mech.* **167**, 353–391 (1986).
18. Arnold, V. I. & Korkina, E. I. *Vest. Mosk. Un. Ta. Ser. 1. Math. Mec.* **3**, 43–46 (1983).
19. Galloway, D. J. & Frisch, U. *Geophys. Astrophys. Fluid Dyn.* **36**, 53–83 (1986).
20. Vishik, M. M. *Geophys. Astrophys. Fluid Dyn.* **48**, 151–167 (1989).
21. Moffatt, H. K. & Proctor, M. R. E. *J. Fluid Mech.* **154**, 493–507 (1985).
22. Proctor, M. R. E. & Weiss, N. O. *Rep. Prog. Phys.* **45**, 1317–1379 (1982).
23. Childress, S. *Phys. Earth planet. Inter.* **20**, 172–180 (1979).

ACKNOWLEDGEMENTS. D.J.G. is grateful for support from the Australian Research Council and from the SERC.

Large air-sea gas fluxes associated with breaking waves

D. W. R. Wallace & C. D. Wirick

Oceanographic and Atmospheric Sciences Division,
Department of Applied Science, Brookhaven National Laboratory, Upton,
New York 11973, USA

THE exchange of gases between the ocean and the atmosphere exerts an important influence on the cycling and global budget of trace gases. Air-sea gas flux is normally parameterized as the product of a gas-transfer velocity and the air-sea concentration difference^{1,2}. Despite suggestions³⁻⁸ that the parameterization might be inappropriate when air-bubble penetration occurs, this 'thin-film' model remains widely accepted⁹ and is employed regularly in regional¹⁰ and global-scale¹¹ models. Here we present a time series of near-surface dissolved O₂ from November to March in coastal waters. The time series is punctuated by sudden large increases in dissolved O₂ associated with surface wave activity. A numerical simulation including air injection by bubbles shows similar behaviour. If the observed O₂ 'events' reflect the air-sea O₂ flux, our results imply that conventional parameterizations might seriously underestimate gas invasion of surface waters under storm conditions. Our study confirms that important questions remain concerning air-sea gas transfer at high wind speeds.

During the Shelf Edge Exchange Processes II (SEEP II) experiment in the Middle Atlantic Bight, during 1988, we deployed two Endeco Model 1133 pulsed polarographic O₂ sensors from a mooring in 38 m of water in a mid-shelf location. The sensors were deployed at depths of 19 m and 34 m, and were accompanied by fluorometers, transmissometers and thermistors. Measurements were made every 68 minutes. Three periods were covered: June-October 1988; November 1988-March 1989; and March-May 1989. Only the winter deployment is discussed here; the summer data will be discussed elsewhere. Sensors were calibrated in sea water equilibrated with three different O₂ partial pressures at three different temperatures. In addition, we did Winkler titrations on water samples collected at the sensor depths before and after each deployment. We estimate the accuracy as ~3%.

We have derived time-series of related meteorological and wave field parameters from three NOAA meteorological buoys located ~100 km north and south of our mooring. Comparison of records indicated that the wind and wave fields were very similar throughout the region and therefore appropriate for our mooring site.

Time series of O₂ saturation and related parameters are presented in Fig. 1. The winter deployment commenced just before the overturning of the water column which occurred at the end of November, as shown by the decrease in the temperature difference and O₂ difference between the upper and lower sensors (Fig. 1c, e). After this, occasional periods of salinity stratification are indicated by the presence of slightly colder water at the 19 m sensor. For long periods the two sensors record near-identical temperatures, but throughout most of the winter the 19 m and 34 m sensors show differing concentrations of dissolved O₂ (Fig. 1c). This difference suggests that the continental shelf water column in this region is rarely well mixed throughout its vertical extent on the timescale of processes affecting dissolved O₂ (days to weeks). This was corroborated by chlorophyll differences between 19 and 34 m (not shown). The O₂ concentrations at 34 m increased steadily throughout the winter, but as the water temperatures decreased gradually in parallel, the saturation remained relatively constant. There is much more variability in the O₂ concentrations recorded at 19 m, with the record being dominated by ~10 instances of rapid increase followed by gradual decreases or 'recoveries'. The apparent systematic O₂ difference between 19 and 34 m is poorly

constrained given our estimated accuracy. Simulations with a two-layer model indicate that the more stable record from the 34 m sensor could result from the smoothing of a fluctuating near-surface signal by vertical mixing.

From January to mid-March, relatively high supersaturations (up to 109%) were recorded at 19 m, and confirmed by more accurate Winkler titrations made during a recovery cruise (12-20 March 1989). It can be seen from Fig. 1 that the sudden increases in O₂ ('O₂ events') do not correspond consistently with wind events, changes in water temperature or the air-sea sensible heat flux. Figure 2 shows that they correspond closely with increases in wave activity as indicated by the significant wave height (Fig. 2a), or, particularly, the amplitude of the orbital velocities at 19 m (Fig. 2b).

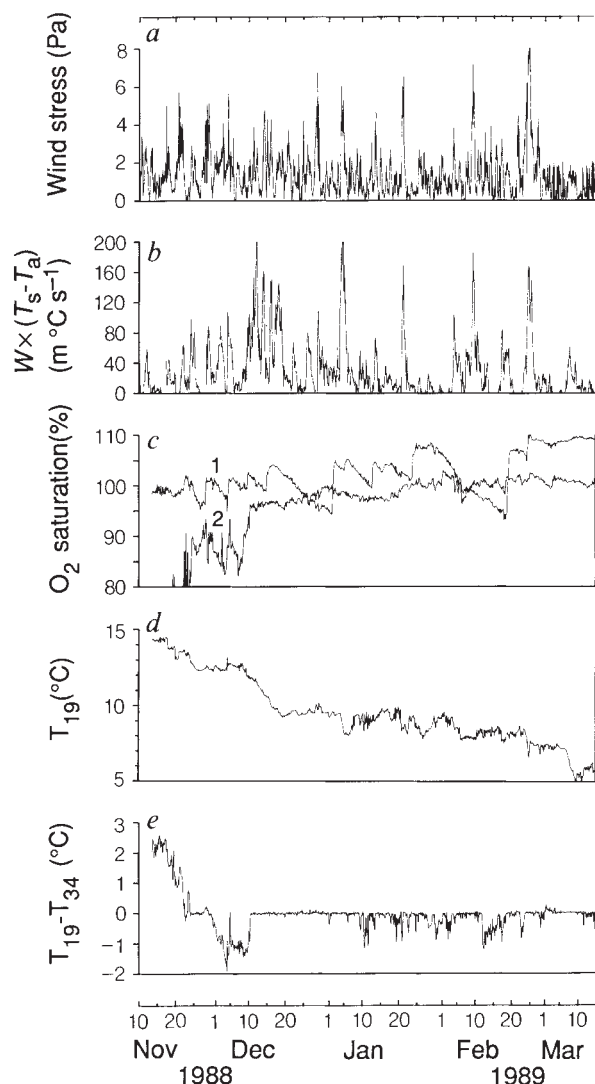


FIG. 1 Time series from November 1988 to March 1989. *a*, Wind stress. *b*, The product of wind speed W and the air-sea temperature difference $T_s - T_a$, which is proportional to the air-sea sensible heat flux. *c*, Dissolved O₂ saturation (%) from two O₂ sensors: (1) at 19 m depth; (2) at 34 m depth. *d*, Water temperature (°C) recorded at 19 m. *e*, Temperature difference (°C) between 19 and 34 m. The O₂ and temperature data were collected at a mooring located at 37° 47.62' N, 74° 44.60' W over the 38 m isobath off the east coast of the United States. Winkler titrations of water samples collected next to the sensors on 20 November 1988 gave 96% (19 m) and 76% (34 m). Corresponding values for 19 March 1989 were 107.5% (19 m) and 99.5% (34 m). The meteorological data were collected at NOAA meteorological buoys 44012, 44019 and CHLV2 located at 38° 30.00' N, 74° 36.00' W, 38° 48.00' N, 74° 36.00' W and 36° 54.00' N, 75° 42.00' W, respectively.

TABLE 1 Measured and predicted air-sea fluxes of O₂

Event	Date	Temperature (°C)	O ₂ (initial) (μM)	O ₂ (final) (μM)	Duration (days)	Wind speed (m s ⁻¹)	Pressure (mbar)	Thin-film flux (μmol m ⁻² d ⁻¹)	Apparent flux (μmol m ⁻² d ⁻¹)
1	28 Nov	12.5	261.0	274.0	0.3	10.6	1,002	2.94 × 10 ⁴	1.40 × 10 ⁶
2	4 Dec	12.7	263.0	273.0	0.3	10.6	1,023	4.03 × 10 ⁴	1.08 × 10 ⁶
3	9 Dec	11.9	273.0	282.0	0.3	7.3	1,019	6.46 × 10 ³	8.33 × 10 ⁵
4	15 Dec	10.5	279.5	297.4	1.4	9.8	1,017	1.50 × 10 ⁴	3.51 × 10 ⁵
5	2 Jan	9.6	273.5	303.0	0.7	6.1	1,016	2.33 × 10 ⁴	1.19 × 10 ⁶
6	5 Jan	8.2	305.0	313.0	0.7	8.4	1,023	-1.32 × 10 ⁴	3.24 × 10 ⁵
7	13 Jan	9.0	291.0	305.0	0.4	6.2	1,016	2.75 × 10 ³	1.01 × 10 ⁶
8	23 Jan	9.3	295.0	317.0	1.0	4.3	1,020	-1.52 × 10 ³	5.94 × 10 ⁵
9	19 Feb	8.1	282.0	319.0	0.6	6.6	1,026	3.40 × 10 ⁴	1.60 × 10 ⁶
10	24 Feb	7.7	313.5	336.0	0.5	18.8	1,015	-1.42 × 10 ⁵	1.33 × 10 ⁶

Comparison, for the 10 events in the O₂ time series (19 m), of the air-sea O₂ fluxes inferred from the measurements with fluxes predicted by a purely thin-film model of air-sea gas exchange using the air-sea concentration difference immediately before the event. Inferred fluxes are calculated assuming that the 19 m sensor is representative of the upper 27 m of the water column. The temperatures, wind speeds and barometric pressures used in the thin-film model calculations are averaged over the duration of the events. A constant salinity of 33 practical salinity units was assumed for calculating O₂ concentrations from the partial pressures measured by the sensors.

The time series were low-pass filtered (with a 40-h cut-off) to remove tidal and diurnal signals, and analysed using standard time-series techniques. Highly significant cross-correlations ($P = 0.001$) were found between hourly changes in oxygen and significant wave height ($r = 0.32$) and wave velocity amplitude ($r = 0.45$); the correlation was highest at lags of < 2 h. No significant cross-correlation was found between oxygen changes and the sensible heat flux, wind stress or hourly temperature changes.

The strong relationship with wave parameters suggested that the sudden O₂ increases were produced by wave breaking and consequent injection of air. The 'recovery' periods would represent periods of net outgassing after the air injection process was switched off, together with O₂ loss due to net community respiration through much of the winter. In the spring, it is possible that net community production of oxygen helps maintain high supersaturation between events. If this interpretation is correct, then our time series presents a striking confirmation of the hypothesized asymmetry of air-sea gas exchange^{3,12} associated with breaking waves. The idea is that gas invades extremely rapidly as a result of pressurization of air bubbles driven several metres below the sea surface (pressure is twice atmospheric at 10 m depth), but evades more slowly via diffusion across the air-sea interface.

We used the O₂ inventory changes at 19 m to estimate an apparent air-sea flux for each of the 10 events (Table 1). Because only two sensors were available, inventory changes were integrated to a depth of 27 m, equidistant between the two sensors. For comparison we present fluxes predicted by the thin-film model using a wind-speed dependence¹³ of the gas transfer velocity appropriate for O₂ (ref. 14). The apparent air-sea flux can be up to three orders of magnitude higher and of opposite sign to the mass flux predicted using the thin-film model. Discrepancies of this nature indicate that a conventional air-sea gas exchange parameterization is inappropriate for gases such as O₂ which can be transported into the ocean by air bubbles.

Such calculations assume that O₂ measured at 19 m responds instantaneously to the air-sea flux: this is equivalent to assuming that the mixed-layer is deeper than 19 m. It should be noted that an asymmetric pattern of sudden O₂ increases and more gradual recoveries could also arise from the migration of a high-O₂ near-surface layer past the sensor (for example because of convective mixing) followed by restratification and respiration. Although this cannot be ruled out without a data set of higher vertical resolution, there is no evidence for it in the time series of temperature, wind or sensible heat flux.

We can, however, assess whether air injection could produce the transient signals observed. Most work on air injection has focused on the mean saturation anomalies of dissolved O₂ in

surface waters^{4,9} and of noble gases in surface and deep waters^{15,16}. Such work defines the long-term average 'dynamic equilibrium' condition, but gives little insight into kinetics or transients under storm conditions. The role of 'extreme events' may be hidden by averaging, by the smoothing caused by ocean mixing, or by inadequate sampling during storms. In the deep ocean, storm-induced mixing might also dilute the instantaneous air injection signal.

Two parameterizations of the air injection flux suitable for real ocean conditions are available. Observations of bubble dynamics and modelling of the associated gas flux by Thorpe⁴ gave

$$F = M_0 C_{\text{sol}} D \left(\frac{2}{9} \frac{g}{D\nu} \right)^{1/3} \exp(0.74 W_{10}) s_1 (s_1 - s) \quad (1)$$

where F is the flux of gas associated with bubbles, $s_1 = (1.5 + 0.5 W_{10})$, s is percentage supersaturation of the gas, W_{10} is wind speed (m s⁻¹) at 10 m height, ν is kinematic viscosity of water, D is molecular diffusivity for the gas in water, g is the acceleration due to gravity, $M_0 = 1.7 \times 10^{-8}$ and C_{sol} is the equilibrium solubility concentration. The flux units are determined by the units of C_{sol} . As yet, there has been no empirical test of whether this model accurately predicts gas fluxes. Alternatively, an empirical gas-cycling model¹⁷, based on monthly sampling of noble gases off Bermuda, gives

$$F = \alpha_i a_{\text{inj}} (f_i + \Gamma f_p) (6 \times 10^{-8}) \left(\frac{W_{10}}{10} \right)^\gamma \quad (2)$$

where α_i is the mole fraction of gas in air, a_{inj} is amplitude of air injection (0.39), $f_i = (1 - f_p)$ is the relative fraction of total (t) and partial (p) bubble trapping ($f_i = 0.07$), $\Gamma = (D_i/D_0)^{2/3}$, where D_i and D_0 are the molecular diffusivities of the gas and of helium at 20 °C, respectively, and γ is an empirically determined exponent (2.2). In this case the flux is given in units of m³ (at standard temperature and pressure) m⁻² s⁻¹.

By coupling these parameterizations with a conventional thin-film gas-exchange model to represent the flux across the sea surface, we have simulated the winter O₂ record for a layer 27 m deep, using hourly data for wind (Fig. 2e) and atmospheric pressure. We also ran a simulation using only the thin-film model. The results (Fig. 2d) indicate a considerable discrepancy between the only two available parameterizations of the air injection flux. The behaviour of the first (model 1 in Fig. 2d) closely resembles our O₂ time-series and suggests that it can be plausibly explained by the injection of oxygen in bubbles. When comparing model results and data, we should remember that the air injection models are forced by the wind and therefore require that gas invasion events correlate with strong wind events, whereas the *in situ* data correlate better with wave

activity. On the continental shelf, the waves may not always be in equilibrium with the wind, so the air injection need not necessarily correlate closely with the wind. The notable disagreement between model 1 and the empirical air-injection model 2 might result from necessarily sparse shipboard sampling during winter for noble gases, and consequent undersampling of 'extreme events', or simply from the limited wind-speed range at Bermuda.

The possible importance of wave-breaking for gas invasion has long been recognized^{12,18} but is not included in thin-film models. A recent test of the wind-speed dependence of the gas

transfer velocity based on SF₆ tracer experiments at high wind speeds¹⁹ did not address this question as it examined gas evasion only. Our data set demonstrates that high-frequency time series of gases such as O₂, N₂ and He, collected from *in situ* instruments or samplers during winter, might be a relatively simple means to examine both gas invasion and evasion. An important goal would be to develop air-injection models in terms of wave parameters that can be remotely sensed by satellite. Although the air-injection effect is likely to be considerably weaker for highly soluble gases such as CO₂ (refs 9, 20), the implications of the large apparent fluxes of Table 1, for dissolved gas studies in general and high-latitude gas transfer in particular, are important. At high latitudes, storms, deep convection and water cooling occur simultaneously, and might result in gas invasion rates that are much higher than currently modelled. □

Received 2 October 1991; accepted 13 March 1992.

1. Liss, P. S. & Slater, P. G. *Nature* **247**, 181–184 (1974).
2. Broecker, W. S. & Peng, T.-H. *Tellus* **26**, 21–35 (1974).
3. Thorpe, S. A. *Phil. Trans. R. Soc. A* **304**, 155–210 (1982).
4. Thorpe, S. A. *Ann. Geophys.* **2**, 53–56 (1984).
5. Memery, L. & Merlivat, L. in *Oceanic Whitecaps* (eds Monahan, E. C. & MacNiocaill, G.) 95–100 (Reidel, Dordrecht, 1984).
6. Merlivat, L. & Memery, L. *J. geophys. Res.* **88**, 707–724 (1983).
7. Lamarre, E. & Melville, W. K. *Nature* **351**, 469–472 (1991).
8. Smith, S. D. & Jones, E. P. *J. geophys. Res.* **90**, 869–875 (1985).
9. Broecker, W. S. *et al. J. geophys. Res.* 10517–10527 (1986).
10. Wallace, D. W. R. & Lazier, J. R. N. *Nature* **332**, 61–63 (1988).
11. Tans, P. P., Fung, I. Y. & Takahashi, T. *Science* **247**, 1431–1438 (1990).
12. Atkinson, L. P. *J. geophys. Res.* **78**, 962–968 (1973).
13. Liss, P. S. & Merlivat, L. in *The Role of Air–Sea Exchange in Geochemical Cycling* (ed. Buat-Menard, P.) 113–127 (Reidel, Dordrecht, 1986).
14. Ledwell, J. J. in *Gas Transfer at Water Surfaces* (eds Brutsaert, W. & Jirka, G. H.) 293–302 (Reidel, Dordrecht, 1984).
15. Craig, H. & Weiss, R. F. *Earth planet. Sci. Lett.* **10**, 289–296 (1971).
16. Craig, H. & Hayward, T. *Science* **235**, 199–202 (1987).
17. Spitzer, W. S. & Jenkins, W. J. *J. mar. Res.* **47**, 169–196 (1989).
18. Kanwisher, J. *Deep-Sea Res.* **10**, 195–207 (1963).
19. Watson, A. J., Upstill-Goddard, R. C. & Liss, P. S. *Nature* **349**, 145–147 (1991).
20. Wolfe, D. K. & Thorpe, S. A. *J. mar. Res.* **49**, 435–466 (1991).
21. Garside, C. & Malone, T. C. *Estuar. Coast. mar. Sci.* **6**, 93–104 (1978).

ACKNOWLEDGEMENTS. The idea to deploy O₂ sensors on the shelf was P. Falkowski's. We thank J. Churchill, C. Flagg, B. D. Johnson, P. Kemp, J. LaRoche, P. Minnett and S. Schwartz for discussions. This work was done under the auspices of the U.S. Department of Energy.

Comparison of two methods for measuring dissolved organic carbon in sea water

Hiroshi Ogawa & Norio Ogura

Department of Environmental Science and Conservation,
Faculty of Agriculture, Tokyo University of Agriculture and Technology,
3-5-8 Saiwaicho, Fuchu-shi, Tokyo 183, Japan

RECENT measurements¹ of the amount of dissolved organic carbon (DOC) in sea water using a high-temperature catalytic oxidation (HTCO) method have generated intense interest because they indicated concentrations several times greater than those obtained from conventional wet chemical oxidation (WCO) methods^{2–4}. As dissolved organic matter in the oceans represents one of the major pools of organic matter in the biosphere^{5,6}, these findings of 'new' DOC have prompted important revisions to models of the oceanic carbon cycle^{7–9}. A satisfactory explanation for the origin of the 'new' DOC, however, which seems to be chemically refractory¹, biologically labile¹⁰ and of high molecular mass¹, has not been forthcoming. Here we present a comparison of measurements of seawater DOC using the HTCO and WCO methods. We obtain fairly good agreement between the two methods, with the HTCO results being considerably lower than those reported previously¹. Our data suggest that the WCO technique may not fail to detect so much DOC as had been previously supposed, and that any carbon that is missed by this method may be in the low- rather than the high-molecular-mass fraction.

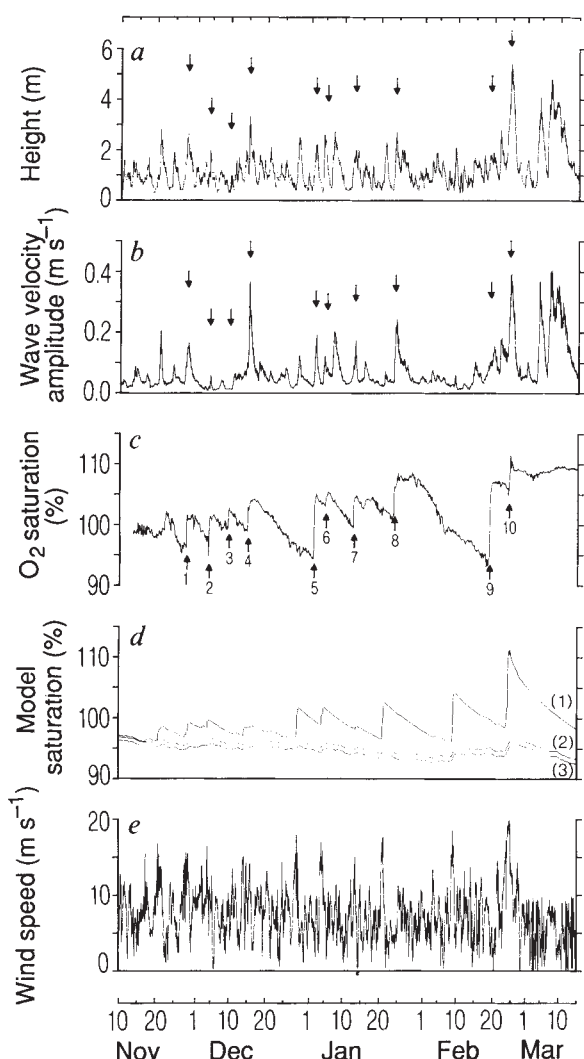


FIG. 2 Observed and modelled time-series. *a*, Significant wave height. *b*, Wave velocity amplitude calculated for a depth of 19 m in 38 m of water; this parameter, calculated from the wave spectrum, refers to the orbital velocities associated with the surface wave field. *c*, The O₂ saturation at 19 m as shown in Fig. 1. *d*, Modelled O₂ saturation for two parameterizations of air injection and a purely thin-film model. *e*, Wind speed. Meteorological and wave data are from the NOAA buoys. The modelled O₂ time series in *d* use equations 1 (line 1) and 2 (line 2) coupled with the thin-film model of gas exchange to simulate the air injection by breaking waves and gas exchange through the air–sea interface. Results from the thin-film model only (no air injection) are given as line 3. The models are forced with the hourly wind (*e*) and atmospheric pressure (not shown) data; simulations commenced on 1 November at 100% saturation. The model output saturation (%) is calculated relative to a constant atmospheric pressure of 1 atm. All simulations incorporate the observed mean water-column cooling rate of 0.073 °C d^{−1} and a net respiration term of 1 μM d^{−1}; the latter term is poorly constrained by field data, but is well within the range possible for winter shelf water²¹.

Estuarine and coastal sea waters were collected in Tokyo Bay during 1990–1991. Oceanic sea waters were sampled in the western North Pacific, off the coast of Honshu (Kanto region), during the KT-91-11 cruise (July, 1991) of *Tansei-maru* (Fig. 1). Immediately or within several hours of collection and storage in the cold (below 4 °C), samples were filtered through pre-combusted glass fibre filters (Whatman GF/C). Bulk DOC determination was made instantly or after the filtrates had been frozen at –20 °C, whereas those for relative molecular mass (M_r) fractionation (see below) were stored at 4 °C after sterilization by addition of 0.1% (v/v) of a saturated solution of HgCl_2 . The preservation techniques used were confirmed by a separate series of experiments using surface samples from Tokyo Bay that showed no significant changes in DOC concentrations between fresh samples and those preserved for about 3 months (our unpublished results).

Dissolved organic matter (DOM) was fractionated according to M_r by ultrafiltration technique¹¹ using Diaflo membranes and an ultrafiltration cell (Amicon). Membranes were YM2 and YM10, which have nominal M_r cut-offs of 1,000 types and 10,000, respectively. Before ultrafiltration, the sample solution was passed through a pre-rinsed Millipore HA membrane filter (0.45- μm pore size). Two high- M_r fractions (>1K and >10K) were then recovered as the concentrates retained on the YM2 and YM10 membranes, respectively. The low- M_r fraction (<1K) was recovered as the filtrate passed through the YM2 membrane. Aliquots of the three M_r fractions obtained for each sample were then subjected to replicate DOC determinations using both WCO and HTOCO, and the results were corrected for the values obtained in running the whole ultrafiltration/determination cycle with double-distilled water treated with persulphate oxida-

TABLE 1 Comparison of the two methods for determination of bulk DOC in the sea waters of the western North Pacific

Station	Depth (m)	Salinity (‰)	*DOC concentration HTOCO	(μM) WCO	WCO/HTCO ratio (%)
Surface					
E12	0	34.1	85.3 $\pm$ 1.5	69.0 $\pm$ 0.0	80.9 $\pm$ 1.4
E15	0	33.6	92.0 $\pm$ 0.7	73.6 $\pm$ 0.5	80.0 $\pm$ 1.1
E20	0	33.3	97.2 $\pm$ 2.0	80.7 $\pm$ 0.5	83.0 $\pm$ 2.2
A1-1	0	32.8	116.8 $\pm$ 1.9	98.4 $\pm$ 1.0	84.3 $\pm$ 2.2
Subsurface					
A1-1	10	34.2	81.2 $\pm$ 0.8	66.8 $\pm$ 1.4	82.3 $\pm$ 2.5
A1-1	50	34.5	65.0 $\pm$ 1.1	49.9 $\pm$ 1.3	76.8 $\pm$ 3.2
A1-1	100	34.5	46.6 $\pm$ 1.3	35.8 $\pm$ 1.3	76.9 $\pm$ 4.9
A1-1	150	34.4	65.8 $\pm$ 0.1	48.8 $\pm$ 1.3	74.2 $\pm$ 2.0
E-12	150	34.4	72.8 $\pm$ 1.6	54.1 $\pm$ 1.4	74.3 $\pm$ 3.6

* Concentrations of organic carbon in the samples filtered through Whatman GF/C glass fibre filter. Results are given as mean $\pm$ s.d. in replicate measurements.

tion. Recovery of DOC during the ultrafiltration process was calculated to be $88 \pm 10\%$ ($n = 33$) from the mass balance as follows: $(\text{DOC}_{<1,000} + \text{DOC}_{>1,000}) / \text{DOC}_{\text{bulk}} \times 100 (\%)$.

After removal of inorganic carbon from the acidified filtrates by purging with nitrogen, the concentrations of organic carbon were determined using WCO with persulphate (the origin method as in ref. 2) and HTOCO done on a commercially available total organic carbon (TOC) analyser (model TOC-5000; Shimadzu) equipped with 20 g of a 3% Pt on Al_2O_3 catalyst. The furnace temperature was maintained at 680 °C and the effluents were passed through a mercury trap (gold wool) to remove mercury vapour when HgCl_2 -preserved samples were injected. (The catalyst and trap were prepared and installed by Shimadzu).

The experimental precisions were typically within $\pm 3\%$ over the range of measurements. Although the blank values obtained for the double-distilled water were usually greater with the TOC analyser (mean $10.5 \pm 1.5 \mu\text{M}$, $n = 23$; 3–4 replicates per sample) than for WCO (mean $3.8 \pm 1.7 \mu\text{M}$, $n = 19$; 2–3 replicates per sample), the causes could not be accurately identified. We have assumed that all blank values were derived only from the double-distilled water, and the system blank was not corrected at all in either method.

The concentrations of bulk DOC (Whatman GF/C filtrates) in the surface waters of Tokyo Bay were $92\text{--}273 \mu\text{M}$ (mean $160 \pm 53 \mu\text{M}$) by HTOCO, and $75\text{--}278 \mu\text{M}$ (mean $159 \pm 60 \mu\text{M}$) by WCO. Those in the bottom waters (15 m at Station 25) were $83\text{--}128 \mu\text{M}$ (mean $103 \pm 18 \mu\text{M}$) by HTOCO, and $75\text{--}121 \mu\text{M}$ (mean $99 \pm 18 \mu\text{M}$) by WCO. The mean ratio of the values obtained by WCO to those of HTOCO was $99 \pm 10\%$, with a good correlation between the two results ($r^2 = 0.946$, $n = 33$). But the maximum values for this eutrophic inshore area are lower than recently reported for oceanic waters¹.

There was close agreement between WCO and HTOCO results for the high- M_r DOM separated from Tokyo Bay surface waters. The results obtained for the >10 K fraction are shown in Fig. 2. The range of *in situ* concentrations found in this fraction was $3\text{--}57 \mu\text{M}$ (as calculated from HTOCO results), accounting for 3–22% of the total DOC ($= \text{DOC}_{<1,000} + \text{DOC}_{>1,000}$). This supports other work that indicates that the high- M_r fraction is generally a minor portion of DOM¹². Similarly, close agreement between the methods were found for the >1K fraction accounting for 32–56% of the total DOC ($\text{DOC}_{\text{WCO}} = 1.04 \times \text{DOC}_{\text{HTCO}} - 21$; $r^2 = 0.992$, $n = 33$; a concentration factor ($\times 10$) was not used). We believe these results are relatively unambiguous in light of the fact that the concentration step inherent in the ultrafiltration procedures reduces the measurement error that can arise in attempting determinations at the miniscule native concentrations of high- M_r fractions.

The two methods were also applied to bulk DOC determinations in nine surface and subsurface samples from the western

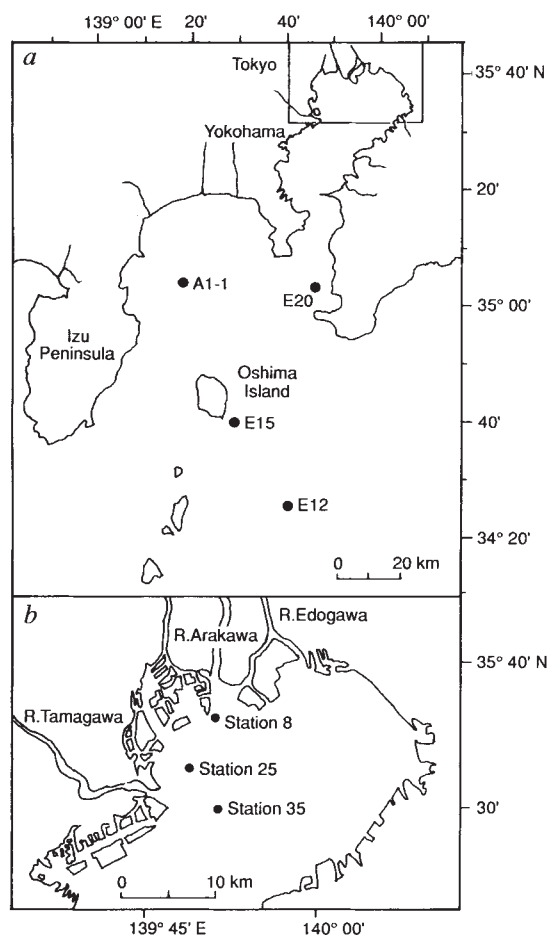


FIG. 1 Seawater sampling sites in western North Pacific (a) and Tokyo Bay (b).

North Pacific. As shown in Table 1, it was apparent that the HTOC results were somewhat higher than those from WCO. Nevertheless, the surface sample HTOC values we obtained using the TOC-5000 (85–117 μM) were considerably lower than those reported earlier using HTOC¹. The differences we found between the two measurements varied only within a range of ~15–25%, and were closely correlated ($r^2 = 0.990$, $n = 9$). These findings agree with those in previous studies on comparisons of seawater DOC methods^{4,13,14}, which strongly supports the conclusion that WCO does not miss an appreciable amount and accurately gives the distribution of DOC in sea water.

The M_r fractionation was examined to pinpoint the methodological differences observed in oceanic waters. The results for surface and sub-surface (150 m) samples from Station E12 (34° 25' N, 139° 40' E; Table 2) reveal differences of only 1–8% between WCO and HTOC results for the two high M_r fractions (>10K and >1K). The greatest differences (33–40%) were found in the low- M_r (<1K) fraction of each sample. The low- M_r fraction seems to be the major portion of DOM, accounting for about 65% of the total DOC (HTCO results). Thus, the methodological differences of 27–28% observed for bulk DOC were primarily derived from the low- M_r fraction, with only a minor contribution from the high- M_r fractions. By contrast, our results for the low- M_r fraction from Tokyo Bay showed a good agreement between the two methods ($\text{DOC}_{\text{WCO}} = 0.940 \times \text{DOC}_{\text{HTCO}} + 0.256$; $r^2 = 0.907$, $n = 35$). Even though WCO gave slightly lower values for this fraction than those from HTOC (as indicated by the slope of the regression curve), these differences did not exceed 20%.

Our results strongly suggest that there is a portion of DOC that is missed in wet oxidation. But it is found in the low- M_r fraction rather than the high- M_r fraction and is much more abundant in open-sea than in coastal waters. Our results at least imply that chemically refractory DOC of high M_r might not exist at all in sea water although DOC undetected by WCO, which is more important in the ocean, can easily be expected to be due to the high- M_r fraction resistant to microbial oxidation^{11,15,16} and probably very old¹⁷.

Alternatively, we cannot conceive of any low- M_r materials resistant to WCO and abundant in oceanic waters. Attention should be focused on finding these substances which are organic

TABLE 2 Comparison of the two methods for determination of organic carbon in different M_r fractions separated from the sea waters of western North Pacific

M_r fraction	*Organic carbon concentration (μM)		WCO/HTCO
	HTCO	WCO	ratio (%)
Surface sample from Station E12			
$M_r > 10\text{K}^\dagger$	4.30 ± 0.26	4.42 ± 0.11	103 ± 9
$M_r > 1\text{K}^\ddagger$	25.9 ± 0.4	23.7 ± 0.5	92 ± 3
$M_r 1\text{K}-10\text{K}^\ddagger$	21.6 ± 0.7	19.3 ± 0.6	89 ± 6
$M_r < 1\text{K}$	44.5 ± 4.3	29.8 ± 0.7	67 ± 8
Initial DOC§	87.5 ± 4.0	63.8 ± 0.8	73 ± 4
Subsurface (150 m) sample from Station E12			
$M_r > 10\text{K}^\dagger$	2.70 ± 0.28	2.67 ± 0.14	99 ± 15
$M_r > 1\text{K}^\ddagger$	26.0 ± 0.5	25.8 ± 0.2	99 ± 3
$M_r 1\text{K}-10\text{K}^\ddagger$	23.3 ± 0.8	23.1 ± 0.3	99 ± 5
$M_r < 1\text{K}$	49.3 ± 2.0	29.4 ± 2.0	60 ± 6
Initial DOC§	70.3 ± 3.2	50.7 ± 1.8	72 ± 6

* Results are given as mean $\pm$ s.d. in replicate DOC measurements.

† Corrected for the concentration factors in the ultrafiltration process.

‡ The difference between the values of $M_r > 1\text{K}$ and $M_r > 10\text{K}$.

§ DOC concentrations in the samples filtered through Millipore HA membranes before ultrafiltration.

species with low-boiling-point behaviours. Volatile organic carbon is generally distinguished from DOC as the portion of TOC that is purgeable (along with inorganic carbon) and it seemed to be only partial¹⁴. But there are volatile species that remain after the decarbonation step and require addition of heat to volatilize¹⁴. Furthermore, there might be volatile intermediate products that could be generated from non-volatile compounds by persulphate oxidation. They could be partly transferred to the vapour phase from the solution within a sealed ampoule under the heating condition used in WCO, and consequently could escape the oxidative degradation and the subsequent detection as CO_2 .

Some support for our hypothesis is found in recent reports of the existence of low- M_r products of photodegradation in open sea, that primarily consisted of carbonyl compounds of relatively low boiling point^{18–20} although these compounds can explain only a small part of the pool missed in WCO. We are pursuing investigations to confirm the existence of these elusive low- M_r materials. □

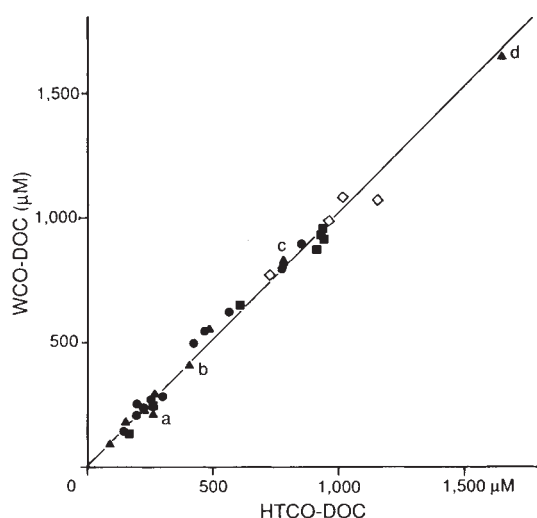


FIG. 2 Correlation between WCO and HTOC results for the high- M_r (>10K) fraction, separated from the surface waters of Tokyo Bay. All samples were concentrated during ultrafiltration by a factor of 20 (except a, b, c and d which were samples of a single sample concentrated by factors of 5, 10, 20 and 40, respectively). Symbols indicate sampling stations: Δ , station 8; $\bullet$, station 25; $\blacksquare$, station 35; $\diamond$, Tamagawa Estuary. $\text{DOC}_{\text{WCO}} = 0.992 \times \text{DOC}_{\text{HTCO}} + 21.5$ ($r^2 = 0.990$, $n = 33$).

Received 22 November 1991; accepted 17 March 1992.

1. Sugimura, Y. & Suzuki, Y. *Mar. Chem.* **24**, 105–131 (1988).
2. Menzel, D. W. & Vaccaro, R. F. *Limnol. Oceanogr.* **9**, 138–142 (1964).
3. Strickland, J. D. H. & Parsons, T. R. in *A Practical Handbook of Seawater Analysis*, Fish. Res. Board Can. Bull. **167**, 310 (1972).
4. Sharp, J. H. *Mar. Chem.* **1**, 211–229 (1973).
5. Williams, P. M. in *Organic Compounds in Aquatic Environments* (eds Faust, S. J. & Hunter, J. V.) 145–164 (Marcel Dekker, New York, 1971).
6. Mopper, K. & Degens, E. T. in *The Global Carbon Cycle* (eds Bolin, B. et al.) 293–316 (Wiley, New York, 1979).
7. Legendre, L. & Gosselin, M. *Limnol. Oceanogr.* **34**, 1374–1380 (1989).
8. Toggweiler, J. R. *Nature* **347**, 122–123 (1990).
9. Bascastow, R. & Maier-Reimer, E. *Glob. biogeochem. Cycles* **5**, 71–85 (1991).
10. Kirchman, D. L., Suzuki, Y., Garside, C. & Ducklow, H. W. *Nature* **352**, 612–614 (1991).
11. Ogura, N. *Mar. Biol.* **24**, 305–312 (1974).
12. Carlson, D. J., Brann, M. L., Mague, T. H. & Mayer, L. M. *Mar. Chem.* **16**, 155–171 (1985).
13. Gershey, R. M., Mackinnon, M. D., Williams, P. J. IeB. & Moore, R. M. *Mar. Chem.* **7**, 289–306 (1979).
14. Mackinnon, M. D. in *Marine Organic Chemistry* (eds Duursma, E. K. & Dawson, R.) 415–443 (Elsevier, Amsterdam, 1981).
15. Barber, R. T. *Nature* **220**, 274–275 (1968).
16. Ogura, N. *Mar. Chem.* **5**, 535–549 (1977).
17. Williams, P. M. & Druffel, E. M. *Nature* **330**, 246–248 (1987).
18. Kieber, D. J., McDaniel, J. & Mopper, K. *Nature* **341**, 637–639 (1989).
19. Kieber, R. J., Zhou, X. & Mopper, K. *Limnol. Oceanogr.* **35**, 1503–1515 (1990).
20. Mopper, K. et al. *Nature* **353**, 60–62 (1991).

ACKNOWLEDGEMENTS. We thank the crewmembers of *Seikai-maru* and the officers of the Harumi Branch of the Bureau of Environmental Protection, Tokyo Metropolitan Government for their help in collecting samples from Tokyo Bay, A. Kuwata of Tokyo University for the distribution of oceanic samples, Y. Morita and staff of the Shimadzu Corporation for technical support and discussion, and O. O'Neill for reading the manuscript.

Dissolved organic carbon in the Atlantic, Southern and Pacific oceans

John H. Martin & Steve E. Fitzwater

Moss Landing Marine Laboratories, Moss Landing, California 95039, USA

THE amount of dissolved organic carbon (DOC) in sea water is controversial^{1,2}. Using a high-temperature catalytic oxidation (HTCO) technique, Sugimura and Suzuki³ reported that surface waters contained 2–4 times as much DOC as that measured previously using wet chemistry and ultraviolet oxidation techniques^{4,5}. They also observed a relationship between DOC content and apparent oxygen utilization suggesting that the consumption of DOC is responsible for oxygen depletion in the deep sea. How to reconcile the apparent differences between these techniques has not been clear. Here we provide independent confirmation of the findings of Sugimura and Suzuki. We collected surface and deep waters from the equatorial Pacific Ocean, the Drake passage and the Atlantic Ocean south of Iceland, and analysed their DOC content using the HTCO methodology³. We found DOC concentrations 2–3 times higher than those measured previously. These results imply that the carbon content of the oceans has previously been underestimated by 10^{12} (1,000 billion) tonnes, and that the new estimated total of 1,800 billion tonnes represents one of the largest carbon reservoirs on Earth⁶. We found no evidence of a cause-and-effect relationship between DOC and apparent oxygen utilization.

We are now confident that, using the new HTCO methodology, we can accurately measure DOC in ocean waters. Basically, the methods we use involve the filtration of water through pre-combusted GF/F glass fibre filters, frozen storage in acid-cleaned, pre-combusted glass bottles with Teflon-lined caps, and analyses using direct injections of 50 μl sea water into a Dohrmann DC-190 HTCO instrument with platinum catalyst

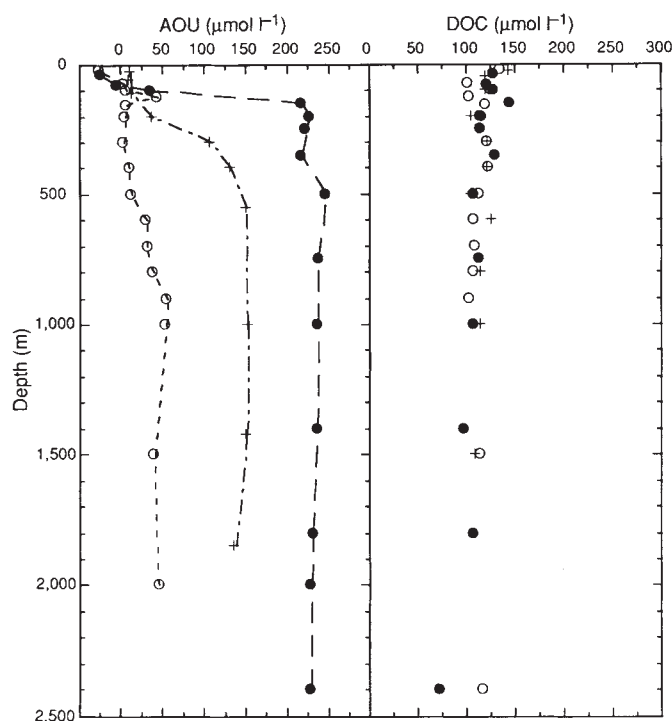


FIG. 1 AOU and DOC profiles from the North Atlantic (○, 59° 27' N; 20° 52' W, 2 July 1989), the Drake Passage (+, 60° 46' S; 63° 26' W, 28 March 1989) and the equatorial Pacific (●, 3° 00' S; 140° 00' W, 7 July 1990).

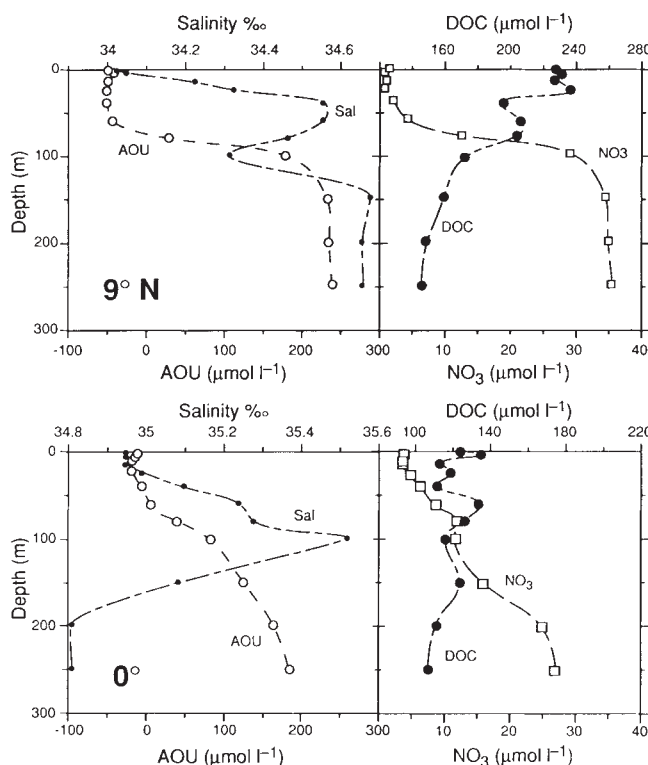


FIG. 2 DOC, NO_3 , salinity and AOU profiles from equatorial Pacific stations at 9° 00' N; 140° 00' W (25 June 1990) and 0° 04' N; 139° 59' W (1 July 1990). During this study, the highest DOC amounts were found in the less salty waters of the North Equatorial Countercurrent (9° N). Saltier, newly upwelled surface waters on the Equator had about the same DOC concentrations as those found at greater depth.

at 680 °C. Further details have been described elsewhere⁷, including intercalibration with Yoshimi Suzuki on Joint Global Ocean Flux Study (JGOFS) North Atlantic samples. Suzuki's fresh samples were analysed immediately after collection, and our frozen aliquots (denoted MLML) were analysed 18 months later; the regression was $\text{DOC}_{\text{MLML}} = -2.3 + 0.956 (\text{DOC}_{\text{Suzuki}})$, $n = 12$, $r = 0.930$, where DOC concentrations are in $\mu\text{mol l}^{-1}$. We also participated in the KGOFS DOC/DON Seattle Workshop^{8,9}; analyses of three replicates from five samples ($n = 15$) resulted in a mean and s.d. of $152 \pm 7.3 \mu\text{mol l}^{-1}$ DOC for North Pacific surface water, 86 ± 5.0 for oxygen minimum water, 112 ± 6.6 for deep water and 135 ± 4.9 for Hawaiian river water.

We compared the DOC content in young (~100 years old)¹⁰ North Atlantic deep waters with those in middle-age (~500 years)¹⁰ southern and old (1,600 years)¹¹ Pacific deep waters. As the subsurface waters age, apparent oxygen utilization (AOU) values steadily increase: the maximum AOU observed south of Iceland was $55 \mu\text{mol l}^{-1}$ compared with $155 \mu\text{mol l}^{-1}$ in the Drake Passage and $245 \mu\text{mol l}^{-1}$ in the equatorial Pacific oxygen minimum (Fig. 1). Nevertheless, despite these great differences in AOU, there were no marked changes in DOC levels; the mean and s.d. for the 17 Iceland samples (depth range = 40–2,400 m) were 114.5 ± 7.4 compared to 117.4 ± 7.1 for the 10 Drake Passage (50–1,500 m) samples and $115.2 \pm 17.4 \mu\text{mol l}^{-1}$ DOC for the 13 equatorial Pacific (60–2,400 m) samples. Furthermore, surface water concentrations were not particularly high at the three stations; that is, 140–150 $\mu\text{mol l}^{-1}$ DOC.

These concentrations are remarkably consistent, and to illustrate that we do obtain significantly different levels from different samples and that our results are 'oceanographically consistent',

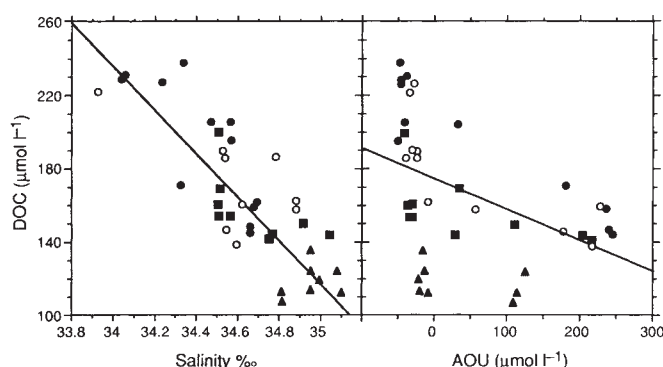


FIG. 3 Plots of DOC against salinity (left) and AOU (right) using near-surface (2–250 m) samples from an equatorial Pacific transect along 140° W (●, 9° N; ○, 6° N; ■, 3° N; ▲, 0°). Regressions were: $\mu\text{mol l}^{-1} \text{DOC} = 3,866 - 106.9 (\text{sal})$; $n=41$, $r=0.828$ and $\mu\text{mol l}^{-1} \text{DOC} = 172.8 - 0.163 (\text{AOU})$; $n=41$, $r=0.489$.

we show data from the ends of a near-surface section across the equatorial Pacific at 140° W (Fig. 2). Although our surface values on the Equator are very close to the average deep amounts, the concentrations are much higher at 9° N. Furthermore, the combined data vary in relation to a hydrographic parameter: as the salinity increases, DOC decreases (Fig. 3). This represents oceanographic consistency. Whether this relationship is coincidental or indicates the introduction of DOC through rainfall¹² or the advection of North Equatorial counter-current low-salinity/high-DOC water from the western Pacific^{3,13} is not known. In contrast to DOC/salinity, little relationship was observed between AOU and DOC (Fig. 3).

If DOC, produced in and sinking away from high-latitude surface waters, is to fuel subsurface oxygen utilization in the global ocean, massive amounts of DOC must be in these regions. In addition to the North Atlantic and Drake Passage stations, we also measured DOC concentrations in iron-rich, productive near-shore surface waters of the Ross Sea (76° 34' S; 167° 29' E, 12 Jan 1990). Although exceptionally high particulate organic carbon concentrations of $74 \mu\text{mol l}^{-1}$ were observed, the maximum surface water DOC value was only $192 \mu\text{mol l}^{-1}$.

Our few high-latitude results indicate that insufficient DOC exists in high-latitude areas to support much consumption in the ocean interior after these waters have sunk away from the surface. Nevertheless, although we find no evidence for a global AOU/DOC relationship, it is very possible that more labile DOC is produced along high-productivity continental margins (for example, Peru) and, with its subsequent consumption, contributes to the intensive oxygen minima occurring there. Vertical mixing must also be considered in regions with large DOC gradients such as those in the western Pacific^{3,13}.

One of the more important aspects of DOC research is the establishment of a reliable estimate for the magnitude of the global ocean DOC pool. For example, Peltzer *et al.* (manuscript in preparation) compare their Atlantic findings with those of Suzuki *et al.*¹³ for the Pacific. Using the new HTOC concentrations they estimate a global DOC pool of $1,630 \times 10^{15} \text{ g C}$ ($10^{15} \text{ g} = 1 \text{ gigatonne}$). This value is 2.3 times greater than that obtained through persulphate or ultraviolet photo-oxidation techniques; that is, 700 Gt of carbon (E. T. Peltzer *et al.*,

manuscript in preparation).

In the above we highlight the consistency of the DOC concentrations we measured in subsurface waters of the Atlantic, Southern and Pacific oceans; that is, about $115 \mu\text{mol l}^{-1}$ DOC or about 1.4 g m^{-3} . Using the world ocean volume¹⁴ for waters deeper than 500 m ($1,173 \times 10^{15} \text{ m}^3$), which is 88% of the total ocean volume, a total of 1,600 Gt C is obtained. Because the remaining 12% (depths less than 500 m) have at least the $1.4 \text{ g m}^{-3} \text{ C}$, our estimate for the global ocean inventory of carbon in the form of DOC is $>1,800 \text{ Gt}$. Considering the limited spatial coverage and that no common locations were used, this estimate and that of Peltzer *et al.* (manuscript in preparation) agree very well.

Although many more analyses are needed to refine these values, there would seem to be little doubt that HTOC-derived global DOC inventories will be at least twice and perhaps three times higher than the 700 Gt obtained with the old methods. Thus the finding of a missing 1,000 Gt DOC pool in the ocean, an amount that exceeds all of the carbon in the atmosphere ($\sim 750 \text{ Gt C}$)⁶ demonstrates the importance of the analytical achievement of Suzuki and the late Yukio Sugimura.

Ogawa and Ogura¹⁵ report that DOC concentrations measured using HTOC were about the same as those determined using conventional wet chemical oxidation. This implies that historic oceanic DOC amounts ($\sim 50\text{--}100 \mu\text{mol l}^{-1}$ DOC) were not underestimated and, conversely, that the high values measured by us and by Suzuki¹⁶ and Tanoue¹⁷ among others are erroneously high. Clearly we disagree. We believe that Suzuki¹⁶ has done a good job of summarizing the current controversy and defending the HTOC methodology for which he is responsible. Tanoue¹⁷ has also reported deep DOC concentrations ($\sim 110 \mu\text{mol l}^{-1}$ DOC) in the northwestern Pacific that are very similar to the deep ocean concentrations we report above. He also fails to find a relationship between AOU and DOC. At a second station Tanoue found HTOC DOC levels ($90\text{--}100 \mu\text{mol l}^{-1}$) that were twice as high as those he measured using a wet chemistry technique. We hope that the reasons for the analytical discrepancies between our findings and those of Ogawa and Ogura¹⁵ can be ascertained in future intercalibration exercises. □

Received 17 January; accepted 17 March 1992.

1. Toggweiler, J. R. *Nature* **334**, 468 (1988).
2. Williams, P. M. & Druffel, E. R. M. *Oceanography* **1**, 14–17 (1988).
3. Sugimura, Y. & Suzuki, Y. *Mar. Chem.* **24**, 105–131 (1988).
4. Holm-Hansen, O., Strickland, J. D. H. & Williams, P. M. *Limnol. Oceanogr.* **11**, 548 (1966).
5. Menzel, D. W. & Ryther, J. H. in *Organic Matter in Natural Waters* (ed. Hood, D. W.) 31–54 (Univ. of Alaska, 1970).
6. Sundquist, E. T. in *The Carbon Cycle and Atmospheric CO₂: Natural Variations Archean to Present* (eds Sundquist, E. T. & Broecker, W. S.) 5–59 (American Geophysical Union, Washington DC, 1985).
7. Fitzwater, S. E. & Martin, J. H. *Mar. Chem.* (in the press).
8. Williams, P. M. *US JGOFS News* **3**, 1 (1991).
9. Hedges, J., Bergamaschi, B. A. & Benner, R. *Mar. Chem.* (in the press).

10. Stuiver, N. M. & Ostlund, H. G. *Radiocarbon* **22**, 1–24 (1980).
11. Ostlund, H. G. & Stuiver, M. *Radiocarbon* **22**, 25–53 (1980).
12. Mopper, K. & Zika, R. G. *Nature* **325**, 246–249 (1987).
13. Suzuki, Y. & Sugimura, Y. *Deep Sea Res.* (in the press).
14. Sverdrup, H. U., Johnson, M. W. & Fleming, R. H. *The Oceans* (Prentice-Hall, Englewood Cliffs, New Jersey, 1941).
15. Ogawa, H. & Ogura, N. *Nature* **355**, 696–698 (1992).
16. Suzuki, Y. *Mar. Chem.* (in the press).
17. Tanoue, E. *Proc. 50th Ann. Meeting Oceanogr. Soc. Jpn.* 142–143 (1991).

ACKNOWLEDGEMENTS. This research was supported by the ONR Ocean Chemistry Program, the NSF Marine Chemistry Program and the Biology and Medicine Section of NSF Polar Programs.

Cyclic formation of debris avalanches at Mount St Augustine volcano

James E. Begét*[‡] & Juergen Kienle^{†‡}

* Department of Geology and Geophysics, † Geophysical Institute and
‡ Alaska Volcano Observatory, University of Alaska, Fairbanks,
Alaska 99775

VOLCANIC debris avalanches have been seen at many volcanoes since the 1980 eruption of Mount St Helens, but typically only one or two avalanche deposits are identified at each eruptive centre, suggesting that catastrophic slope failures are rare or even unique events in the lifetime of a volcano¹⁻⁴. Here we present a series of radiocarbon dates from volcanic deposits showing that the summit edifice of Mount St Augustine, a 1,220-m-high active volcano on Augustine Island in the Cook Inlet area of south-central Alaska, has repeatedly collapsed and regenerated, averaging 150–200 years per cycle, during the past 2,000 years. The unprecedented frequency of summit edifice failure was made possible by sustained lava effusion rates over 10 times greater than is typical of plate-margin volcanoes.

Mount St Augustine has historically been the most active volcano in the eastern Aleutian arc of Alaska. Pyroclastic and dome-building eruptions occurred in 1812, 1883, 1935, 1963–64, 1976 and 1986⁵⁻⁷. The historic eruption of 1883 produced the Burr Point debris avalanche⁸, which covered at least 20 km² on the north flank of Augustine Island, extending the coastline by 2 km and travelling more than 4 km into Cook Inlet (Figs 1 and 2). Displacement of sea water by the debris avalanche produced a 20-m-high tsunami at the island. Wave run-up as high as 10 m occurred over 100 km from the volcano⁹. Dome growth since the 1883 debris avalanche has buried the amphitheatre crater left by it and created a new, oversteepened summit dome complex, raising concerns that a future eruption might culminate in another debris avalanche and tsunami⁸⁻¹⁰.

Studies of long-term eruption history are a useful technique for identifying patterns of volcanic behaviour and estimating typical repose times between eruptions^{8,11-14}. There are no written records of the behaviour of Mount St Augustine before 1812, but sea cliffs around the perimeter of the roughly 100-km² island expose a series of discrete and overlapping debris avalanches which are locally separated by alluvium, lahars, pyroclastic flows, tephra layers and thin buried peat layers.

The well-exposed debris avalanches have characteristic features seen at similar deposits at other volcanoes, including irregular, hummocky topography, lithic heterogeneity, poor sorting, and discrete zones of shattered fresh rock and highly altered rock debris separated by sheared avalanche matrix^{1,3,15}. The borders of the avalanches are delimited by semi-continuous lateral levees, 5–15 m high, which together with cross-sectional exposures in sea cliffs allow the number and dimensions of the deposits to be accurately mapped (Fig. 2). We determined the frequency of past debris avalanches at Mount St Augustine and evaluated the likelihood of a future avalanche by making a detailed stratigraphical and geochronological study of these prehistoric volcanic deposits.

Previous studies of Mount St Augustine indicated that the volcano was very young, and that its earliest eruptions occurred in latest Pleistocene time about 10,000–15,000 years ago^{7,16}. But accelerator mass spectrometry (AMS) radiocarbon dating of small charcoal fragments found in the oldest fragmental volcanic rocks on the island shows that they are more than 40,000 years old (Table 1). These deposits of basaltic hyaloclastites and silicic pyroclastic flows and tephra occur with glacial deposits as high as 330 m above sea level¹⁵, indicating that the volcano had grown at least to this height before the last glaciation^{16,17}. One debris

avalanche exposed on the east side of Augustine Island may also be Late Pleistocene or Early Holocene in age.

The central dome complex at Mount St Augustine is surrounded by very recent, overlapping debris avalanches which cover 90% of the area of Augustine Island, and are named for local geographic features (Fig. 2). We have obtained more than 30 radiocarbon and AMS dates on soils, peat and buried wood found above and below these avalanche deposits, and on intercalated tephra layers and pyroclastic flow deposits which contain charred wood. These radiometric dates closely bracket the age of the avalanche deposits (Table 1).

The frequency of edifice collapse can be assessed by converting the radiocarbon dates on debris avalanches to calendar years, using recent dendrochronological calibrations of the

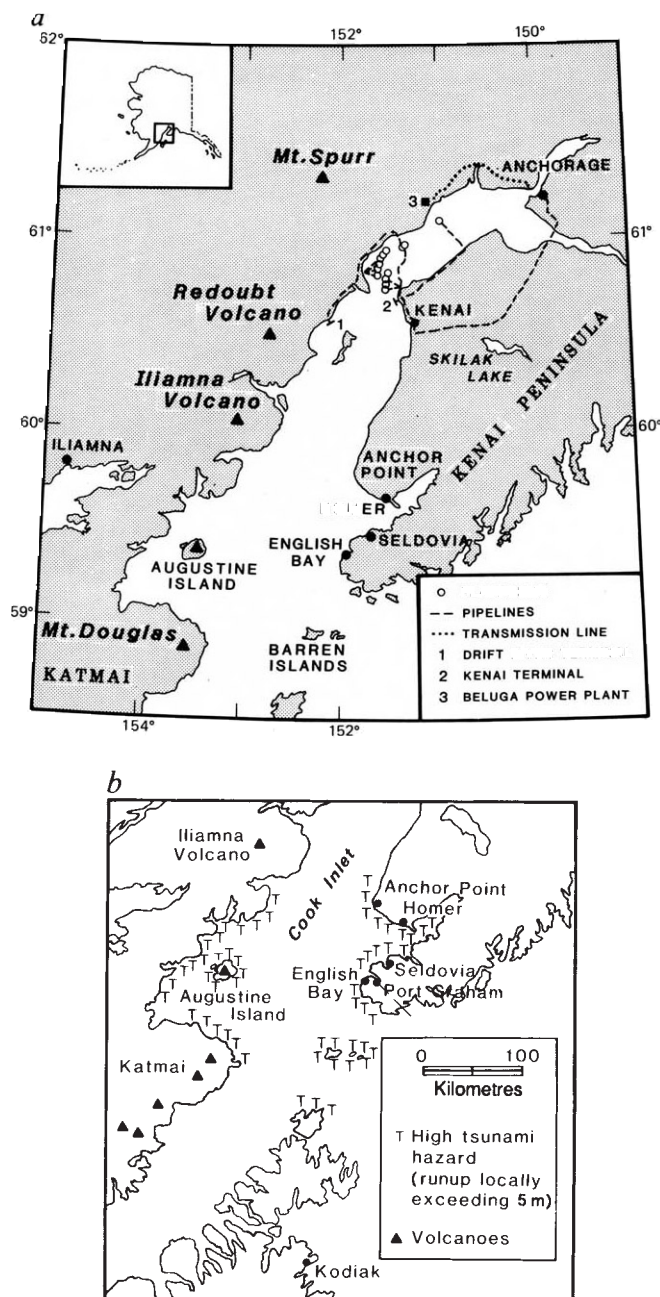


FIG. 1 a, Location of Mount St Augustine and Augustine Island in lower Cook Inlet, Alaska and other major volcanoes in the Cook Inlet area. Also shown are the locations of cities and towns, offshore oil platforms, coastal oil terminals, powerplants and power transmission lines in coastal areas. b, Areas of especially high tsunami hazard.

TABLE 1 Stratigraphy of radiocarbon dates, tephra layers and volcaniclastic deposits at Mount St Augustine

Radiometric age	Calendar year (AD)	Laboratory number	Stratigraphic context	Reference
Burr Point debris avalanche (AD 1883)				
<185	—	I-14,922	U, P	This study
≤285	—	I-14,923	U, S	This study
140 ± 60	—	B-24773	U, S	This study
170 ± 70	—	B-24779	U, S	This study
195 ± 115	—	—	U, S	5
205 ± 90	1664 (1638–1700)	I-14,924	U, S	This study
Rocky Point debris avalanche				
280 ± 100	1642 (1480–1670)	B-24781	U, P	This study
330 ± 145	1521, 1590, 1623 (1430–1670)	—	U, L, S	5
West Island debris avalanche				
Tephra layer 'B'				
367 ± 55	1488 (1442–1529)	—	I, B	8
410 ± 50	1446 (1434–1491)	B-24777	U, L, S	This study
450 ± 80	1437 (1412–1487)	B-28536	U, L, S	This study
470 ± 140	1432 (1390–1520)	—	U, L, S	5
490 ± 70	1426 (1398–1444)	B-24776	L, S	This study
Grouse Point debris avalanche				
730 ± 60	1278 (1257–1284)	B-28537	U, S	This study
Tephra layer 'M'				
770 ± 50	1261 (1218–1279)	—	—	—
Southeast Beach debris avalanche				
Lagoon debris avalanche				
800 ± 120	1230, 1243, 1256 (1150–1280)	—	U, L, S	5
860 ± 80	1191 (1148–1260)	B-24780	U, L, S	This study
1,000 ± 100	1004, 1008, 1019 (980–1130)	B-24780	U, L, S	This study
Tephra layer 'C'				
Southwest pyroclastic fan				
Long Beach debris avalanche				
1,130 ± 50	895, 922, 939 (865–982)	B-28535	C, I, U	This study
1,195 ± 120	780, 804, 823, 827, 841, 857 (670–980)	—	U, L, S	5
1,200 ± 140	780, 790, 802, 843, 853 (660–990)	—	U, L, S	5
1,290 ± 80	686, 754, 757 (652–779)	B-28539	C, I, U	This study
1,400 ± 50	643 (602–663)	B-28534	C, I, U	This study
1,420 ± 60	640 (576–658)	B-28538	U, L, S	This study
South Point debris avalanche				
Tephra layer 'H'				
1,470 ± 160	584, 587, 597 (410–680)	—	U, L, S	This study
1,500 ± 155	558 (410–660)	—	U, L, S	This study
1,610 ± 70*	426 (382–538)	ETH 3826	U, L, P	This study
Northeast Point debris avalanche				
Tephra layer 'I'				
1,830 ± 80	134, 152, 169, 200, 211 (72–257)	B-24775	U, L, S	This study
Southeast Point debris avalanche				
East Point debris avalanche				
Tephra layer 'G'				
Yellow cliffs debris avalanche				
Late Pleistocene volcanic deposits				
≥ 39,890*	—	ETH 7166	I, C	This study
≥ 40,440*	—	ETH 7167	I, C	This study

Calendar ages and range (at 1 s.d.) based on dendrological calibration from ref. 18.

U, upper age limit; L, lower age limit; I, incorporated in deposit; C, burned alder branch; P, peat; S, soil.

* AMS date.

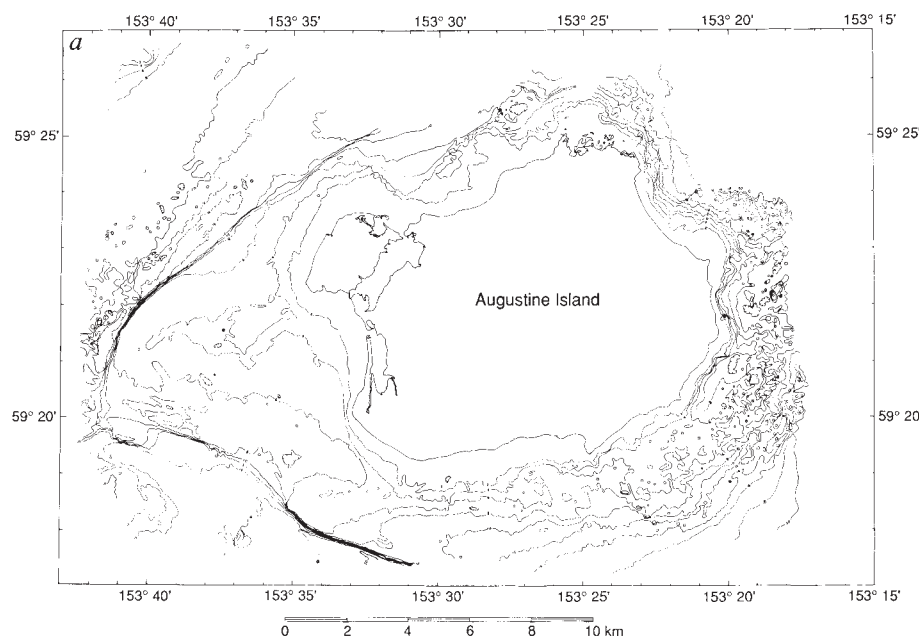
radiocarbon method¹⁸. Altogether, at least 11 discrete large debris avalanche deposits formed during the past 1,800–2,000 years can be identified in sea cliffs on the flanks of Mount St Augustine. Seven of these, at Burr Point, Rocky Point, West Island, Grouse Point, Southeast Fan, Long Beach and the Lagoon, are less than 1,000 years old, four occurred in the last 800 years, and three occurred in the past 500 years, with the most recent in 1883. The avalanche deposits are typically 5–30 m thick in seacliff exposures, and cover areas similar to the historic 1883 avalanche. Thus, a geological record lasting two millennia records repeated debris avalanche activity with an average recurrence interval of about 150–200 years.

All 11 avalanches mapped by this study entered the sea, and their submarine extents can be determined from the irregular hummocky topography found extending 3–6 km offshore on

almost all flanks of the island as seen on a new, detailed bathymetric survey (Fig. 2). Because the travel distance and overall mobility of debris avalanches is in part a function of the source area elevation^{19–21}, the similar extent of these deposits suggests they originated from about the same elevation and that the Mount St Augustine central dome complex regenerated after each cycle of collapse.

The repeated emplacement of debris avalanches requires that unusually high lava effusion rates have been maintained at Mount St Augustine for the past two millennia. Following the 1883 debris avalanche, a horseshoe-shaped crater similar to that created in 1980 at Mount St Helens occupied the summit of Mount St Augustine²² and had a volume of $2.5 \times 10^8 \text{ m}^3$, whereas the avalanche deposit has a greater volume of about $3.0\text{--}3.5 \times 10^8 \text{ m}^3$ which is due to fragmentation and expansion of the dome

FIG. 2 *a*, Bathymetric map showing hummocky topography associated with offshore extent of debris avalanches. *b*, Distribution of debris avalanches on Augustine Island and hummocky terrain offshore showing their lateral extent. BP, Burr Point Avalanche; RP, Rocky Point avalanche; GP, Grouse Point avalanche; WI, West Island avalanche; L, Lagoon avalanche; LB, Long Beach avalanche; SP, South Point avalanche; RB, Rocky Beach avalanche; SE, Southeast avalanche; YC, Yellow Cliffs avalanche; NEP, Northeast Point avalanche; EP, East Point avalanche.



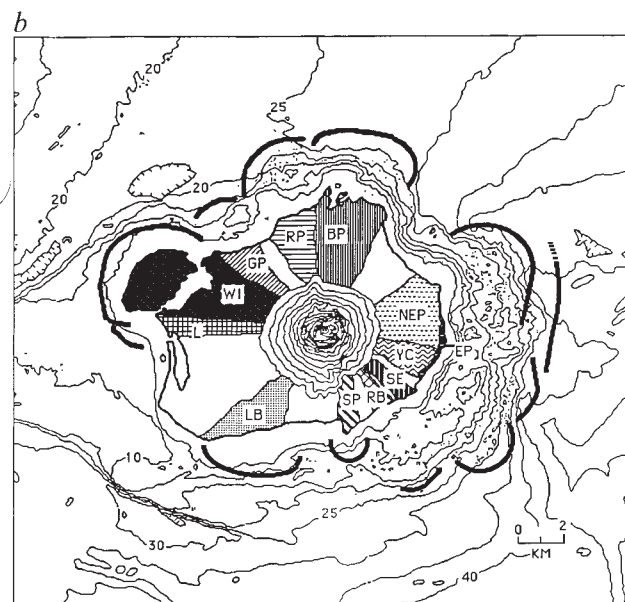
debris during transit downslope⁸. The present summit dome complex was produced by intermittent eruptions in 1883–84, 1935, 1964, 1976 and 1986. It has almost completely buried the 1883 crater and is about 800 m in diameter and 520 m high, with an estimated volume of $2.6 \times 10^8 \text{ m}^3$ (Fig. 3).

We have evaluated long-term eruption rates by estimating the cumulative volume of the dated prehistoric avalanche deposits. Lava and dome-building eruptions during just the last century reflect an average effusion rate of about $2.5 \times 10^6 \text{ m}^3 \text{ yr}^{-1}$ (exclusive of small tephra falls), whereas the series of 11 avalanches deposits indicates average effusion rates of about $1\text{--}3 \times 10^6 \text{ m}^3 \text{ yr}^{-1}$. This eruption rate, although much less than found at active hotspot volcanoes like Mauna Loa and Kilauea in Hawaii^{23,24}, is as much as 10–100 times greater than for typical calc-alkaline arcs and volcanoes^{25–27}, indicating that during the past 2,000 years Mount St Augustine has sustained one of the highest known rates of volcanic output attributable to subduction-zone-related magmatism²⁸.

The high eruption rates of silicic domes and lava flows at Mount St Augustine repeatedly produced an oversteepened, unstable volcanic cone which then collapsed. Volcanoes are usually thought to consist almost entirely of lava flows and pyroclastic rocks, but Mount St Augustine shows that in some cases volcanic edifices may consist largely of debris avalanche deposits. This mode of behaviour may be short-lived, but at Mount St Augustine it has produced a radial volcanic platform 15 km in diameter and tens to hundreds of metres thick (Fig. 2).

More than 70 tsunamis of volcanic origin worldwide have caused devastation and at least 50,000 fatalities during the past 200 years^{4,29–31}. In North America, apart from Alaska, there are few volcanoes that lie near enough to the sea to generate tsunamis during eruptions. The past pattern of behaviour at Mount St Augustine indicates there is a high likelihood of another debris avalanche occurring in the near future. The current sea shore lies only 3–7 km from the summit dome complex, whereas previous avalanches have typically travelled 7–11 km from the summit (Fig. 2). The 1883 tsunami produced a small amount of damage to boats and houses in scattered coastal villages, but today about 250,000 people live near Cook Inlet, and offshore oil platforms, oil storage and tanker facilities, fishing and recreational marinas are in coastal areas which might be affected by the recurrence of such a tsunami (Fig. 1).

The tsunami hazard may be exacerbated by the regional tectonic setting of Mount St Augustine. The volcano lies in the



part of Alaska affected by great subduction zone earthquakes of M8 or greater. A small M5 earthquake contributed to the collapse of Mount St Helens in 1980, and sharp earthquakes probably caused the 1792 Unzen volcanic debris avalanche in Japan which produced a tsunami and perhaps 15,000 fatalities⁴. Future large earthquakes could easily weaken the summit dome complex and induce edifice collapse at Mount St Augustine, quite independently of future eruptive behaviour.

Concentric north-facing 10–20-m-high scarps appeared around the southern part of the 1986 dome and the summit during the 1986 eruption of Mount St Augustine (Fig. 3), raising concerns that deep-seated fault planes were propagating through the volcano, but the volcano has been seismically quiet and the summit area stable since the end of the 1986 eruption. An island-wide geodetic deformation net involving 30 EDM (electronic distance measurement) lines and a short-distance EDM net at the volcano summit was established in 1988 to monitor future creep and deformation of the summit edifice.

Although it is at present quiescent, the high frequency of

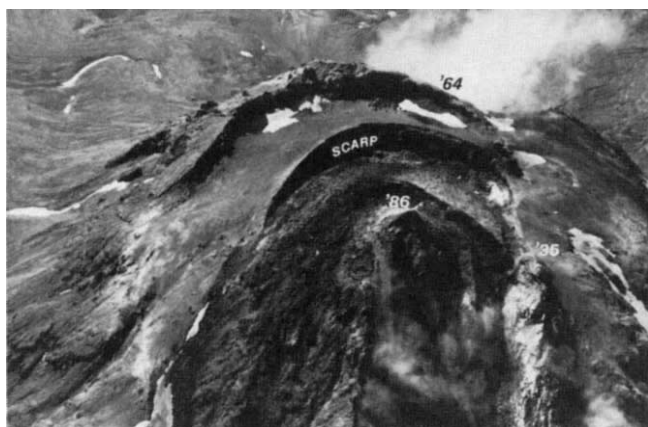


FIG. 3 The summit of Mount St Augustine, Alaska as viewed from the north, showing the location of the 1935, 1964 and 1986 domes. The prominent, concentric, north-facing scarps, which are 10–20 m high, developed during the 1986 eruption. The coastline of Augustine Island and the waters of Cook Inlet lie only 4–6 km to the north of the summit.

historic eruptions and the evidence of an average 150–200 year recurrence interval between large debris avalanches during the past 2,000 years raises concerns about the future stability of Mount St Augustine. This concern is compounded by the complete reconstruction of an oversteepened summit edifice by four major dome-building eruptions since the 1883 debris avalanche. Any future dome eruptions will contribute to additional growth and oversteepening or deformation of the summit region. If Mount St Augustine continues to follow the pattern that it has maintained for the past 2,000 years, it is likely that another debris avalanche will occur sometime in the next century, which in turn will produce a tsunami by displacing the waters of Cook Inlet. □

Received 25 November 1991; accepted 25 February 1992.

1. Uli, T. *J. Volcan. geotherm. Res.* **18**, 135–150 (1983).
2. Siebert, L. *J. Volcan. geotherm. Res.* **22**, 163–197 (1984).
3. Uli, T., Yamamoto, H. & Suzuki-Kamata, K. *J. Volcan. geotherm. Res.* **29**, 231–243 (1986).
4. Siebert, L., Glicken, H. & Uli, T. *Bull. volcanol.* **49**, 435–459 (1987).
5. Swanson, S. E. & Kienle, J. *J. geophys. Res.* **93**, 4500–4520 (1988).
6. Kienle, J. & Swanson, S. E. *Volcanic Hazards from Future Eruptions of Augustine Volcano, Alaska*, (Geophysical Institute, Univ. of Alaska, Fairbanks, 1985).
7. Kienle, J. in *Volcanoes of North America* (eds Wood, C. A. & Kienle, J.) 79–80 (Cambridge Univ. Press, 1990).
8. Siebert, L., Glicken, H. & Kienle, J. *Nat. geogr. Res.* **5**, 232–249 (1989).
9. Kienle, J., Kowalik, Z. & Murty, T. S. *Science* **236**, 1442–1447 (1987).
10. Kienle, J., Davies, J. N., Miller, T. P. & Yount, M. E. *EOS* **67**, 580–582 (1986).
11. Crandell, D. R., Mullineaux, D. R. & Rubin, M. *Science* **187**, 438–441 (1975).
12. Begét, J. E. *Science* **215**, 1389–1390 (1982).
13. Tilling, R. I. *et al. Science* **224**, 747–749 (1984).
14. Tilling, R. I. (ed.) *Volcanic Hazards* (American Geophysical Union, Washington DC, 1989).
15. Glicken, H. *US Geol. Surv. Prof. Paper* **1488** (1991).
16. Johnston, D. A. *US Geol. Surv. Circ.* **B804**, 78–80 (1979).
17. Detterman, R. L. *US Geol. Surv. Map* GQ-1068 (1973).
18. Stuiver, M. & Becker, B. *Radiocarbon* **28**, 863–910 (1986).
19. Hsu, K. *Geol. Soc. Am. Bull.* **86**, 129–140 (1975).
20. Eppler, D., Fink, J. & Fletcher, R. *J. geophys. Res.* **92**, 3623–3633 (1987).
21. Melosh, H. J. in *Debris Flows/Avalanches: Process, Recognition and Mitigation* (eds Costa, J. E. & Wieczorek, G. F.) 41–50 (Geological Society of America, Boulder, Colorado, 1987).
22. Becker, G. F. *US Geol. Surv. Ann. Rep.* **18**, 28–58 (1898).
23. Holcomb, R. T. *US Geol. Surv. Prof. Paper* **1350**, 243–260 (1987).
24. Lockwood, J. P. & Lipman, P. W. *US Geol. Surv. Prof. Paper* **1350**, 509–536 (1987).
25. McBirney, A. A. *Rev. Earth planet. Sci.* **6**, 437–456 (1978).
26. Sugimura, A., Matsuda, T., Chinzei, K. & Nakamura, K. *Bull. volcanol.* **26**, 125–140 (1963).
27. Civetta, L., Cornette, Y., Gillot, P. Y. & Orsi, G. *Bull. volcanol.* **50**, 47–57 (1988).
28. Crisp, J. J. *Volcan. geotherm. Res.* **20**, 177–211 (1984).
29. Latter, J. H. *Bull. volcanol.* **44**, 467–490 (1981).
30. Simkin, T. & Fiske, R. S. *Krakatau 1883* (Smithsonian Institution, Washington DC, 1983).
31. Blong, R. J. *Volcanic Hazards* (Academic, Orlando, 1984).

ACKNOWLEDGEMENTS. We acknowledge the cooperation and support of our colleagues in the Alaska Volcano Observatory and the University of Alaska's Institute of Marine Science and of L. Siebert of the Smithsonian Institution. We dedicate this contribution to H. Glicken, lost in the 1991 eruption of Unzen Volcano, Japan. This work has been funded by the US NSF (J.E.B. and J.K.), and by support from the Alaska Volcano Observatory, jointly operated by the Geophysical Institute of the University of Alaska, the State of Alaska Division of Geological and Geophysical Surveys, and the US Geological Survey.

First discovery of monotremes in South America

Rosendo Pascual*, Michael Archer†, Edgardo Ortiz Jaureguizar*, José L. Prado*, Henk Godthelp† & Suzanne J. Hand†

* Departamento Científico Paleontología Vertebrados, Facultad de Ciencias Naturales y Museo Universidad Nacional de La Plata, Paseo del Bosque, 1900 La Plata, Argentina

† School of Biological Science, University of New South Wales, PO Box 1, Kensington, New South Wales 2033, Australia

UNTIL now, the egg-laying monotremes were only known from the Australian continent, where they have lived since the early Cretaceous period to the present¹. Here we report the first monotreme from outside the Australian continent, an ornithorhynchid, from sediments of late early Palaeocene age in Patagonia, southern Argentina. This discovery demonstrates the Gondwanan nature of monotremes and supports the hypothesis that the Patagonian Terrane of southern South America had a biotic history distinct from that of the rest of the continent.

The South American monotreme is represented by an isolated upper right second molar (Fig. 1, M²). The crown morphology is monotreme-like and resembles the upper second molars of Oligo-Miocene species of the ornithorhynchid genus *Obdurodon*^{2,3} in having two parallel anteroposteriorly compressed, occlusally concave, V-shaped blade systems (triakididrepanons^{2,4}) terminated buccally by four small cusps. At the lingual end of the midvalley of the tooth, between the bases of the two transverse blade systems, there is a low ridge (or cuspule) that probably occluded with the apical margin of the principal buccal cusp of the 'talonid' of the lower molars. At the buccal end of the midvalley, there are two additional minor cusps that contribute to the ornithorhynchid-like multi-cuspidate buccal margin of the crown. As in the M² of species of *Obdurodon*, the Argentinian tooth has posterior and anterior cingula, no buccal or lingual cingula and four anteroposteriorly compressed roots. In contrast to species of *Obdurodon*, the tooth is larger, has an elevated posterior cingulum, a more buccal position of the posterobuccal cuspule, a swollen basal portion of the anterolingual cusp and extensive occlusal wear. The Argentinian animal, although only known from an upper molar, appears to have been larger than *Steropodon galmani*, which is known from a portion of a lower jaw. A detailed comparison with all other monotremes will be presented elsewhere^{3,4}.

Enamel ultrastructure of the Argentinian tooth resembles that reported for *Obdurodon*⁵. The middle enamel shows a Pattern 2 enamel-packing pattern⁶; tubules are present and there is a distinct inter-row sheet and inter-prism component present. Prisms are generally ovoid and ~3 µm in diameter. This combination of features exists in *Obdurodon* and some marsupials but not in placentals or multituberculates (C. Gilkeson, personal communication). The inability to examine ultrastructural features on etched surfaces prohibited a definitive examination for marsupial versus monotreme features. Gross crown morphology is almost indistinguishable from the pattern of ornithorhynchid monotremes but is unlike that of any marsupial or placental mammals.

The mammal-bearing deposit ('Banco Negro Inferior', BNI) occurs at 45° 30' S; 67° 11' W, Golfo de San Jorge, Chubut province (central Patagonia), Argentina. The BNI deposit occurs above the marine Salamanca Formation⁷, whose uppermost part has been determined to be Upper Danian on the basis of foraminiferal biostratigraphy^{8,9}. On the basis of palaeomagnetic data, the BNI sediments have been determined to be between 63.2 and 61.8 Myr BP (before present; ref. 10). Mammals from this level are most similar to those from the Tiupampian Local Fauna of Bolivia¹¹, both assemblages having been referred to

the Tiupampian Land Mammal Age of South America^{10,12}. The Argentinian platypus tooth was found on the surface of an eroded exposure of BNI sediments (further locality data are available from R.P.).

The BNI Local Fauna include five crocodiles (two mesosuchians, the crocodylid *Necrosuchus ionensis* Simpson, 1937, the alligatorids *Eocaiman* sp. and a probable representative of the North American genus *Allognathosuchus*), four species of chelid turtles, a species of the gondwanatherian multituberculate genus *Sudamerica*¹³, at least two primitive 'condylarths', including a possible arctocyonid and a mioclaenid, and a representative of a derived but as yet undetermined group of placental mammals. This faunal assemblage, lithogenesis and the occurrence of mangrove pollen suggest that the deposits accumulated in a freshwater facies of a quiet coastal lagoon or mangrove swamp¹⁴.

The absence of monotremes and gondwanatherians from the first known and relatively contemporaneous northern Bolivian Tiupampian faunal assemblages^{10,11} appears to support the hypothesis, based on geobiotic evidence, of a distinct southern South American biota¹⁴⁻¹⁹. Geotectonic studies suggest that southern South America should be distinguished as the Patagonian Terrane²⁰.

A cladistic analysis of 17 taxon cladograms (extant taxa) led to the hypothesis that the biotas of southern South America, Australia, Tasmania, New Guinea, New Caledonia and New Zealand constitute a monophyletic group²¹. But, considering that the oldest known monotreme has been recorded from early Cretaceous sediments in Australia¹, it is surprising that this group has not hitherto been recorded from the late Cretaceous Los Alamitos Local Fauna of Patagonia^{22,23}.

Among Patagonian Tiupampian mammals, monotremes and the autapomorphic gondwanatherian multituberculate genus *Sudamerica*^{13,24} appear to be survivors of lineages present in

South America since the early Cretaceous. The period of time between the early Cretaceous and the earliest Palaeocene was the first of two significant periods of isolation for South America²⁵. Its end was heralded by the arrival of groups such as marsupials prior to the beginning of Tiupampian time. The second significant period of South American isolation extended from the beginning of Tiupampian time to the Pliocene, when the rise of the Bolivar Geosyncline precipitated the Great American Biotic Interchange²⁶. Among mammals, only monotremes and highly autapomorphic multituberculates survived the first period of isolation and persisted into the beginning of the second—but possibly only in Patagonia because they are unknown from Tiupampian¹¹ or younger deposits further to the north²⁷.

In Australia, nothing is known about monotremes between the time of their first record in the early Cretaceous (*Steropodon galmani*, possibly an ornithorhynchid) and the late Oligocene to middle Miocene (*Obdurodon*, ornithorhynchid), and then again until the early Pliocene to Holocene (*Ornithorhynchus*, ornithorhynchid)^{1-3,28}. Tachyglossids are unknown prior to the late Miocene (*Megalibgwilia*²⁵), although DNA hybridization experiments suggest they diverged from ornithorhynchids between the late Cretaceous and early Tertiary²⁹. The South American taxon is thus the only early Tertiary monotreme known.

Evidence supports an interordinal relationship to relatively plesiomorphic 'eupantotheres' such as dryolestoids¹⁻³ rather than more derived groups such as peramurids^{1,30}, with a probable origin sometime in the Jurassic. The South American monotreme is most similar among monotremes to the Oligo-Miocene species of *Obdurodon*, although upper molars are as yet unknown for *Steropodon galmani*. For this reason, we tentatively refer it to the Family Ornithorhynchidae. But because it shares a conservative molar morphology exhibited by all toothed monotremes which spans the early Cretaceous to the middle Miocene, it contributes little information that would clarify interordinal relationships of the group. It is remarkable that the basic molar pattern has changed so little from the early Cretaceous to the middle Miocene, a uniquely long period of stasis in molar morphology for any group of mammals.

Australia did not completely separate from East Antarctica until between the middle Eocene and early Oligocene (45–38 Myr) and South America maintained at least an archipelagic connection to West Antarctica until the late Oligocene (~35 Myr)³¹. Recent phylogenetic studies have suggested that at least one order of marsupials (Ameghino's Microbiotheria) or its descendants spanned the three unified continents before the early Tertiary^{32,33}. It is now clear that a comparable distribution pattern characterized monotremes.

The distribution of ornithorhynchids in South America and Australia reflects similar palaeobiogeographic histories for the two island continents¹⁹. Both started out as parts of the nuclear supercontinent Gondwana, presumably sharing a regional biota, before drifting away in a northerly direction. While South America was intermittently isolated during its northerly migration until it became united with the North American plate^{12,25}, its southern tip (the Patagonian Terrane) trailing in cold temperature regions, Australia has been more effectively isolated during its still ongoing drift into subtropical regions. Their different post-separation histories resulted in a progressively different balance of the Gondwanan groups both once shared, such as ceratodontids among lungfish¹⁹ and monotremes among mammals. □

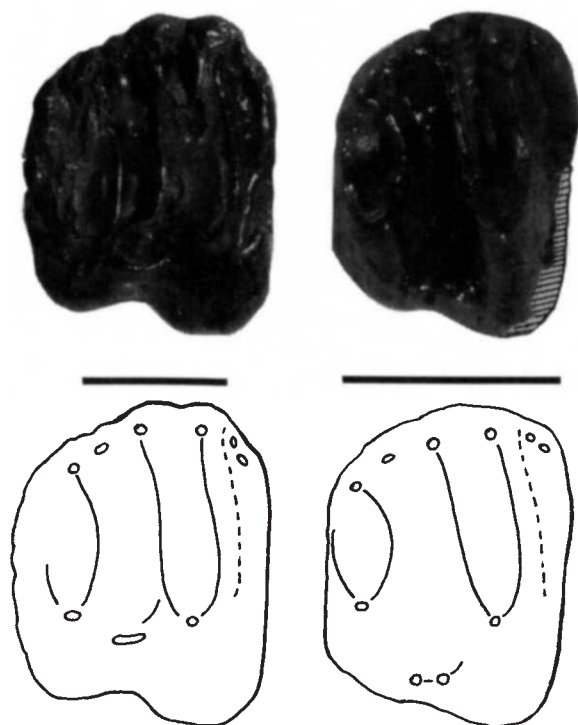


FIG. 1 The worn Banco Negro Inferior ornithorhynchid right M² from Patagonia (left) compared with a less worn but slightly damaged right M² of *Obdurodon* sp. from Oligo-Miocene deposits in Riversleigh, northwestern Queensland. Anterior is to the right; the curved row of small cusplules is buccal; lingual is down. Diagrams (below) compare distinctive primary cusp and blade relationships (Scale bar, 5 mm).

Received 28 October 1991; accepted 14 February 1992.

1. Archer, M., Flannery, T. F., Ritchie, A. & Molnar, R. E. *Nature* **318**, 363–366 (1985).
2. Archer, M., Murray, P., Hand, S. J. & Godthelp, H. in *Mammal Phylogeny* (eds Szalay, F., Novacek, M. & McKenna, M.) (Springer, New York, in the press).
3. Archer, M., Jenkins, F. A. Jr., Hand, S. J., Godthelp, H. & Murray, P. *Aust. Zool. Rev.* (in the press).
4. Every, R. G. *A New Terminology for Mammalian Teeth Founded on the Phenomenon of Thegosis* (Pegasus, Christchurch, 1972).

5. Lester, K. S. & Archer, M. *Anat. Embryol.* **174**, 145–151 (1986).
6. Borde, A. thesis, Univ. London (1964).
7. Legarreta, L. A., Uliana, M. A. & Torres, M. A. in *Actas II Simp. sobre el Terciario de Chile: Procesos y Productos de Depositación en el Ambiente Andino* 135–176 Univ. de Concepción, Chile, 1991).
8. Méndez, I. A. *Rev. Asoc. Geol. Arg.* **21**, 127–160 (1966).
9. Maziuk, V. *Rev. Mus. La Plata, nueva serie, sección Paleont.* **5**, 197–258 (1967).
10. Pascual, R. & Ortiz-Jaureguizar, E. In *Facies y Fósiles de Bolivia* Vol. 2 (ed. Suárez-Soruco, R.) 3–4 (Rev. Tecn. Yac. Petrol. Bolivia, 1991).
11. Marshall, L. G. & de Muizon, C. *Nat. Geog. Res.* **4**, 23–55 (1988).
12. Pascual, R. & Ortiz-Jaureguizar, E. *J. hum. Evol.* **19**, 22–60 (1990).
13. Kraus, D. E. & Bonaparte, J. F. *J. vert. Paleont.* **10**, 31A (1990).
14. Petriella, B. T. P. & Archangelsky, S. in *Actas I Congr. Arg. Paleont. Bioestrat.* **2**, 257–270 (Tucumán, Argentina, 1975).
15. Zinsmeister, W. J. in *Historical Biogeography, Plate Tectonic and the Changing Environments* (eds J. Gray & A. J. Boucot) 349–355 (Proc. 37th Ann. Biol. Colloq., Oregon University Press, 1979).
16. Zinsmeister, W. J. *J. Paleont.* **36**, 84–102 (1982).
17. Vilas, J. F. & Valencio, D. in *Acta V. Congreso Latinoam. Geol.* **3**, 743–758 (1982).
18. Case, J. A. *Geol. Soc. Am.* **169**, 523–530 (1988).
19. Pascual, R. & Bondesio, P. in *Actas VI Congr. Geol. Arg.* **1**, 565–577 (Buenos Aires, Argentina, 1975).
20. Ramos, V. A. *Am. Scient.* **77**, 444–450 (1989).
21. Crisci, J. V., Cigliano, M. M., Morrone, J. J. & Roig-Juñent, S. *Syst. Zool.* **40**, 152–171 (1991).
22. Bonaparte, J. F. in *Symp. Evol. Vert. Mesoz. Actas 2* (ed. Bonaparte, J. F.) 63–95 (IV Congr. Arg. Paleont. & Bioestrat., Mendoza, Argentina, 1986).
23. Bonaparte, J. F. (ed.) *Rev. Mus. Arg. Cs. Nat., Paleontología III*(3), 103–178 (1987).
24. Scillato-Yané, G. J. & Pascual, R. *Ameghiniana* **21**, 173–76 (1985).
25. Pascual, R. & Ortiz-Jaureguizar, E. in *Fossil Mammals and Evolution* (eds Forstén, A., Fortelius, M. & Werdelin, L.) (*Acta Zool. Fen.* **191**, in the press).
26. Stehlí, F. G. & Webb, S. D. (eds) *The Great American Biotic Interchange*, XVII and 532 (Plenum, 1985).
27. Marshall, L. G., Hoffstetter, R. & Pascual, R. *Paleovertebrata* Montpellier, Mém. Extr. 1983, 1–93 (1983).
28. Woodburne, M. O. & Tedford, R. H. *Am. Mus. Novit.* **2588**, 1–11 (1975).
29. Westerman, M. & Edwards, D. *Abstr. Symp. Monotreme Biol.* (ed. Augee, M.) 27 (Roy. Zool. Soc. New South Wales, Sydney, 1991).
30. Kielan-Jaworowska, Z., Crompton, A. W. & Jenkins, F. A. Jr *Nature* **326**, 871–873 (1987).
31. Veevers, J. J. (ed.) *Phanerozoic Earth History of Australia* (Clarendon, Oxford, 1984).
32. Szalay, F. in *Carnivorous Marsupials* (ed. Archer, M.) 621–640 (Roy. Zool. Soc. New South Wales, Sydney, 1982).
33. Aplin, K. & Archer, M. in *Possums and Opossums: Studies in Evolution* (ed. Archer, M.) XV–LXXII (Surrey Beatty and the Roy. Zool. Soc. NSW, Sydney, 1987).
34. Pascual, R. *et al. Aust. Zool. Rev.* (in the press).
35. Griffiths, M., Wells, R. T. & Barrie, D. J. *Aust. Mammal.* **14**, 87–101 (1991).

ACKNOWLEDGEMENTS. We thank Amoco Oil Company and Astra for assistance; L. Legarreta and M. A. Uliana for unpublished data; R. Arnett for photographs. We also thank the CONICET, Argentina, M. McKenna, the Australian Research Council, the Department of Arts, Sport, Environment, Tourism and Territories, the Australian Geographic Society, the Riversleigh Society, the Royal Zoological Society of New South Wales for assistance in funding this research.

Antiquity of clonal salamander lineages revealed by mitochondrial DNA

C. M. Spolsky, C. A. Phillips* & T. Uzzell

Department of Ecology, Ethology and Evolution, University of Illinois, Urbana, Illinois 61801, USA

THE existence of clonally reproducing vertebrates has often served as a foil in attempts to explain the near-ubiquity of sexual reproduction in eukaryotes, but the absence of recombination, with its attendant limitation of new genotypes to those produced through mutations, restricts the adaptive ability of clonal organisms^{1–3}. It has been argued, therefore, that clonal vertebrate taxa have short lifespans^{4–14}. Variation in mitochondrial DNA (mtDNA) within clonal populations is interpreted instead as reflecting multiple, although limited, independent hybridization events^{8–13,15}. On the basis of an analysis of an average of 373 nucleotide pairs, we report here that the mtDNA of clonal, hybrid, gynogenetic mole salamanders (*Ambystoma*, Ambystomatidae) differs by 5% or more from mtDNA of their closest possible sexual relatives (*A. jeffersonianum*, *A. laterale* and *A. texanum*). Assuming usual rates of mtDNA divergence, these lineages have persisted for about 5 million years, far longer than estimated for other clonal vertebrate populations. The low mtDNA variability in the clonal lineages suggests that they have undergone population reductions during the Pleistocene.

Triploid, gynogenetically reproducing lineages of mole salamanders are abundant in the northeastern United States and

southeastern Canada¹⁶. They arose by hybridization between *A. jeffersonianum* and *A. laterale*^{16–19} and, except for irregularities in occasional individuals, contain either two complete chromosome sets from *A. laterale* and one from *A. jeffersonianum* (*A. tremblayi*), or two chromosome sets from *A. jeffersonianum* and one from *A. laterale* (*A. platineum*). These ongoing lineages are all female, and require sperm from males of a sexual host species for activation of their eggs (gynogenesis). Oogenesis both in *A. platineum* and in *A. tremblayi* includes a prediplotene and presumably premeiotic endoduplication of chromosomes and formation of 42 pseudobivalents consisting of paired sister chromosomes followed by an apparently normal meiosis²⁰, resulting usually in triploid ova genetically identical to maternal somatic cells. Such oogenesis, by maintaining the somatic genetic constitution unchanged from generation to generation, permits these triploid lineages to persist. The observation of electrophoretic variation at several protein-encoding loci has, nevertheless, led to the interpretation that hybrid lineages are evolutionarily very unstable^{21,22}.

Our examination of mtDNA from these triploid hybrid lineages of *Ambystoma* and the sexual species known^{16,19} or speculated²³ to have contributed genetic material to them provides no evidence for either recent or multiple origins from independent hybridization events for these clonal lineages, even in areas where the parental species occur close together or are sympatric. Mitochondrially, the lineages not only are stable, but are ancient, and have persisted far longer than has been deduced for other clonally reproducing vertebrates.

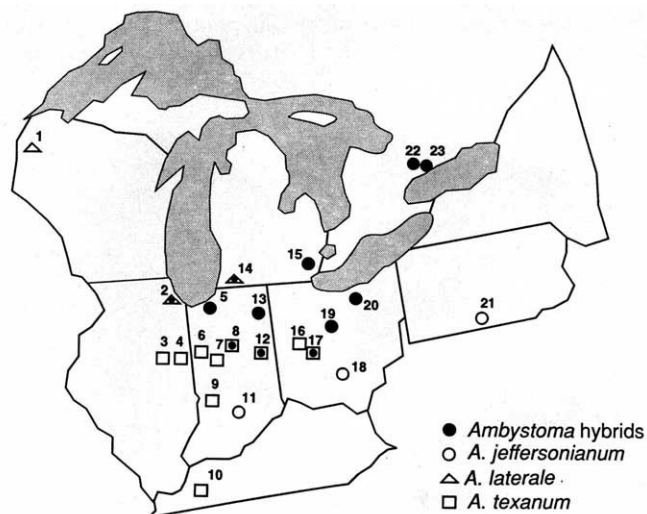


FIG. 1 Northeastern United States and southeastern Canada showing sources of specimens examined. Locality numbers correspond to numbers on the map. Numbers of animals from each locality are in parentheses. 1, Wisconsin, Burnett Co.: *A. laterale* (1); 2, Illinois, Cook Co.: *A. laterale* (2), *A. platineum* or *A. tremblayi* (9), diploid hybrid (1), tetraploid hybrid (1); 3, Illinois, Champaign Co.: *A. texanum* (1); 4, Illinois, Vermilion Co.: *A. texanum* (1); 5, Indiana, Porter Co.: triploid hybrid (1); 6, Indiana, Fountain Co.: *A. texanum* (1); 7, Indiana, Montgomery Co.: *A. texanum* (1); 8, Indiana, Clinton Co.: *A. texanum* (2), *A. platineum* (14); 9, Indiana, Clay Co.: *A. texanum* (1); 10, Kentucky, Christian Co.: *A. texanum* (1); 11, Indiana, Brown Co.: *A. jeffersonianum* (1); 12, Indiana, Delaware Co.: *A. texanum* (1), *A. platineum* (3), diploid hybrid (1); 13, Indiana, Allen Co.: triploid hybrids (2); 14, Michigan, Cass Co.: *A. laterale* (1), *A. tremblayi* (2); 15, Michigan, Oakland Co.: *A. tremblayi* (1); 16, Ohio, Hardin Co.: *A. texanum* (1); 17, Ohio, Brown Co.: *A. texanum* (1), *A. platineum* (1), tetraploid hybrid (1); 18, Ohio, Franklin Co.: *A. jeffersonianum* (1); 19, Ohio, Wyandot Co.: triploid hybrid (1); 20, Ohio, Lorain Co.: *A. platineum* (2); 21, Pennsylvania, Franklin Co.: *A. jeffersonianum* (1); 22, Ontario, Halton Co.: *A. platineum* (3), *A. tremblayi* (1); Ontario, Peel Co.: *A. platineum* (4).

* Present address: Department of Biology, Washington University, St Louis, Missouri 63130, USA.

TABLE 1 Number of individuals, number of localities, number of haplotypes, and interhaplotype divergence (mean and range) in clonal and sexual *Ambystoma*

	Sample size	Number of localities	Number of haplotypes	Interhaplotype divergence (%) (mean and range)
<i>A. jeffersonianum</i>	3	3	3	1.3 (1.0–1.6)
<i>A. laterale</i>	4	3	3	0.4 (0.3–0.6)
<i>A. texanum</i>	11	9	8	1.2 (0.2–2.3)
Clonal lineages	48	12	2	0.1 (0.1)

Samples of mtDNA were obtained from 66 salamanders including 48 clonal hybrids from 12 localities, three *A. jeffersonianum* from three localities, four *A. laterale* from three localities, and 11 *A. texanum* from 10 localities (Fig. 1). Ploidy of hybrids was determined by erythrocyte size (measured from dried blood smears) or DNA content (measured by flow cytometry); our sample contained 44 triploid *A. platineum* and *A. tremblayi*, two tetraploid and two diploid hybrids. The mtDNAs were cleaved with 18 hexanucleotide-recognizing restriction enzymes, producing between 54 and 74 fragments, equivalent to between 324 and 444 nucleotide pairs per taxon. The mtDNA in the 48 clonal hybrid individuals was virtually invariant (Table 1); only two haplotypes were detected, differing by one *EcoRI* site (0.1% sequence divergence). The southern haplotype (A), with three *EcoRI* sites, is found in the 37 individuals (whether *A. platineum*, *A. tremblayi*, or hybrids of other ploidy) from Illinois, Indiana, and Ohio; a northern haplotype (B), lacking one of the *EcoRI* sites, is present in the animals from Michigan and Ontario.

Much greater variation was detected within each of the three sexual species: a different haplotype was found at virtually each locality: only *A. texanum* from Fountain and Montgomery counties in Indiana, which are very close, shared the same haplotype. Within *A. jeffersonianum* and *A. texanum*, average divergence between haplotypes was 10 times higher than between haplotypes of the clonal hybrids (Table 1). Although the number of nucleotide differences among *A. laterale* haplotypes was low (nearly as low as between the two clonal haplotypes), mtDNA haplotype diversity in *A. laterale* is considerably greater than in the clonal lineages: a different haplotype was present in the one or two individuals examined from each population of *A. laterale*, a striking contrast to the two haplotypes seen among 48 individuals from 12 widely scattered populations of clonal salamanders. The low intraspecific divergence values in *A. laterale* compared to *A. jeffersonianum* and *A. texanum* may reflect more limited geographical sampling but probably results from greater population reduction during the Wisconsin glaciation^{16,22}.

Although the two mtDNA haplotypes in the hybrid lineages are very similar to each other, they are markedly divergent from mtDNAs in the three sexual species (Table 2). Consistent with earlier observations²³, mtDNA of the clonal lineages appears least divergent from mtDNA of *A. texanum*. The greater similarity may indicate more recent common ancestry of *A. texanum* mtDNA and that of the clonal lineages, but without an outgroup as a reference point to determine the rooting of phylogenetic trees, derivation of hybrids remains unresolved. Comparison of trees constructed with and without assuming equal rates of evolution²⁴ suggests different rates among the four basic mtDNA groups (Fig. 2). An *F*-test²⁵ based on comparison of residual sums of squares for such trees rejects the hypothesis that the two trees provide equally good estimates of branch lengths ($F = 6.59$; d.f. = 10, 45; $P < 0.01$). The tree constructed without assuming equal branch lengths provides a significantly better estimate of branch lengths; this implies that rates of evolution

TABLE 2 Average mtDNA divergences between species of *Ambystoma* associated with unisexual clonal lineages

	J	L	T	3n
<i>A. jeffersonianum</i> (J)	74	7.8	9.2	7.9
<i>A. laterale</i> (L)	0.255	67	8.5	8.5
<i>A. texanum</i> (T)	0.201	0.260	58	5.2
Clonal hybrids (3n)	0.296	0.247	0.431	67

For each species, numbers along the diagonal (in bold) are the average number of restriction fragments generated per species by the 18 enzymes used. Numbers above the diagonal are average interspecies per cent sequence divergences calculated from the proportion of shared restriction fragments; numbers below the diagonal are the average interspecies proportion of shared restriction fragments. Restriction enzymes used were *AvaI*, *BanII*, *BclI*, *BglII*, *EcoO109I*, *EcoRI*, *EcoRV*, *HaeII*, *HpaI*, *NdeI*, *NsiI*, *PvuII*, *SacI*, *SpeI*, *SstII*, *StuI*, *StyI* and *XbaI*.

in the four taxa are discordant. The rates in some *A. texanum* and in clonal lineages are slower than in *A. jeffersonianum* and *A. laterale*; these slower rates may spuriously link the former two. Preliminary sequence data (our unpublished results) support this interpretation.

The large divergence of mtDNA of the clonal forms from that of any sexual *Ambystoma* species (5.2–8.5%) suggests a long independent existence of the clonal lineages, and provides a maximum age for them. Although the clonal lineages themselves could have arisen more recently, we have found no mtDNA haplotype resembling that in clonal *Ambystoma* (compare with ref. 23). On the basis of our estimate of average rates of mtDNA sequence divergence between pairs of amphibian species, 1% per 10^6 yr (our unpublished results; compare with refs 26, 27),

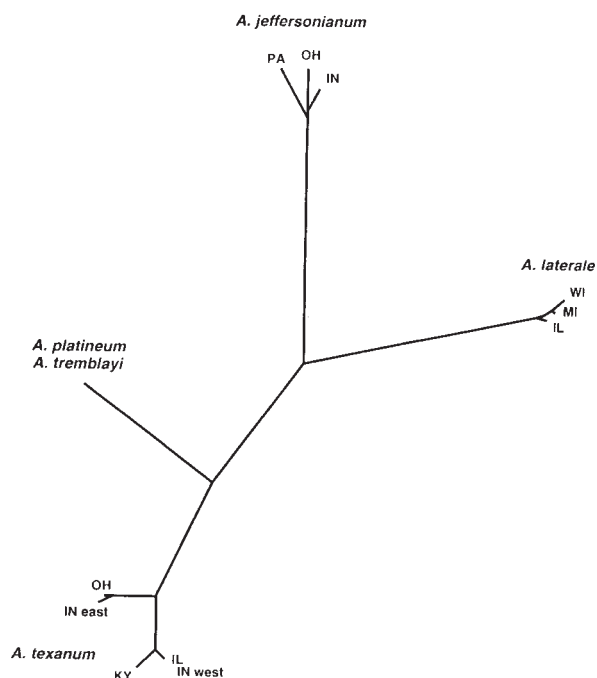


FIG. 2 Phylogenetic tree based on sequence divergences among mtDNA haplotypes. Tree was constructed using FITCH²⁴; this program does not assume equality in sister branch lengths. Branch lengths shown are proportional to calculated distances. For this analysis, the following very similar haplotypes were combined: (1) *A. texanum* from Champaign Co., Illinois, and from the four western localities in Indiana (Montgomery, Clinton, Clay, and Fountain Cos); (2) *A. texanum* from the two Hardin Co., Ohio localities; and (3) two haplotypes of clonal hybrids. PA, Pennsylvania; OH, Ohio; IN, Indiana; WI, Wisconsin; MI, Michigan; IL, Illinois; KY, Kentucky.

the observed divergence (at least 5.2%) may have accumulated over a span of about 5×10^6 yr. This contrasts with most clonal species of hybrid origin, in which divergences from the maternal parental species are usually less than 1%. The deduced maximum age of other clonal vertebrate lineages is thus considerably less than that of clonal *Ambystoma*.

The long-term persistence of clonal species of *Ambystoma* provides independent evidence that the germ-line modifications that allow their persistence from generation to generation are also effective in the long term. Given the apparent great age of these clonal lineages, the very low diversity (in both number of haplotypes and number of nucleotide differences) is surprising. The uniformity of mtDNA suggests that the mtDNA of all extant clonal lineages is derived from a single sexual female. If two or more sexual females with identical mtDNAs had been involved in the original hybridizations some 5×10^6 yr ago, the mtDNAs in their clonal descendants would subsequently have followed separate evolutionary trajectories and therefore would have diverged much more than 0.1%.

The very low clonal diversity could be a consequence of either selection or of stochastic processes resulting in severe real or effective restrictions in the number of reproducing clonal individuals. Selection (compare with ref. 28) seems an unlikely explanation for three reasons: (1) because both nuclear and mitochondrial genomes in a clonal organism would face the same selection process, there should be no nuclear genetic variability; (2) selection should not act on silent substitutions, so restriction site changes involving them should still occur; (3) selection requires both migration and competitive replacement, population by population, of all existing haplotypes by the favoured haplotype, a slow process. Alternatively, severe geographic population restriction requires only subsequent emigration and expansion. Population restriction is plausible given the biogeography of the *A. jeffersonianum* complex: *A. platineum* and *A. tremblayi* occur almost exclusively in areas covered by the Wisconsin glaciation; these taxa may have survived only in small refugia during each of the Pleistocene glaciations and repopulated glaciated areas during each interglacial period.

The diversity in nuclear genotypes of clonal lineages may reflect infrequent replacement of nuclear genomes in clonal species by genomes from males of the sexual host species. That *A. platineum* and *A. tremblayi* both have both clonal haplotypes suggests that such genome replacement through male hosts may have occurred, although replacement has not been demonstrated in natural populations of *Ambystoma*, and must be infrequent²⁹. Even rare replacement, however, may be important in maintaining the nuclear variability present in clonal *Ambystoma* and thus in increasing long-term survival of clonal salamander lineages. □

24. Felsenstein, J. in *Cladistics: Perspectives in the Reconstruction of Evolutionary Biology* (eds Duncan, T. & Steussy, T. F.) 169–191 (Columbia Univ. Press, New York, 1984).
25. Sheldon, F. H. *Molec. Biol. Evol.* **4**, 56–69 (1987).
26. Brown, W. M., George, M. Jr & Wilson, A. C. *Proc. natn. Acad. Sci. U.S.A.* **76**, 1967–1971 (1979).
27. Shields, G. F. & Wilson, A. C. *J. Molec. Evol.* **24**, 212–217 (1987).
28. Brower, A. V. Z. & Boyce, T. M. *Evolution* **45**, 1281–1286 (1991).
29. Spolsky, C. M., Phillips, C. A. & T. Uzzell. *Evolution* (in the press).

ACKNOWLEDGEMENTS. We thank G. Rabb, M. McKnight, P. Rabenold, J. List, D. Rice, D. Sever, A. Sih, R. Humbert, P. Jacobsen, S. Minton and G. Casper for help with collection of specimens and F. Gill for space at the Academy of Natural Sciences of Philadelphia. This work was supported by an NSF grant to C.M.S.

Ancestry of unisexual salamanders

S. Blair Hedges, James P. Bogart* & Linda R. Maxson

Institute of Molecular Evolutionary Genetics, Department of Biology, Pennsylvania State University, Pennsylvania 16802, USA

* Department of Zoology, University of Guelph, Guelph, Ontario N1G 2W1, Canada

IN eastern North America there are populations of all-female salamanders that incorporate the nuclear genomes of two or three of four sympatric bisexual species. The hybrids can be diploid, triploid, tetraploid or pentaploid, and 18 different combinations have been reported. All hybrids require sperm from a sympatric male of one of the bisexual species to reproduce, but the sperm may or may not be incorporated in the egg. Some of the hybrids are believed to represent separate, clonal species, but little is known of the origin of this hybrid complex. Vertebrate mitochondrial DNA is inherited maternally, allowing identification of the female parent that gave rise to hybrid lineages. A portion of the cytochrome *b* gene was sequenced from diploid and triploid hybrids that represent combinations of all four species. Nearly all hybrids had a similar mitochondrial genome sequence, independent of nuclear genome composition and ploidy, and the sequence was distinct from that of any of the four bisexual species. The hybrids maintain a mitochondrial lineage that has evolved independently of their nuclear genome and represent the most ancient known unisexual vertebrate lineage.

To determine the maternal ancestry of the unisexual salamander (genus *Ambystoma*) hybrids, we sequenced part of the mitochondrial (mt) cytochrome *b* gene from 46 salamanders, including diploid and triploid hybrid combinations of all four bisexual species (*A. jeffersonianum*, *A. laterale*, *A. texanum*, *A. tigrinum*; ref. 1). The cytochrome *b* gene evolves sufficiently rapidly to allow determination of the relationships among closely related populations and species of vertebrates². We extracted and purified DNA from salamanders with known chromosome numbers and allozyme genotypes^{3–7}. To compare the bisexual species with these hybrids, we chose samples of *laterale*, *jeffersonianum*, *texanum* and *tigrinum* from widely separated populations, to examine the range of sequence differentiation within these species. The hybrid samples included diploid *laterale* × *texanum*, where the two species occur sympatrically, and diploid *laterale* × *jeffersonianum*, where the two species occur parapatrically. The triploid combinations *laterale* × 2-*jeffersonianum* (LJJ), 2-*laterale* × *jeffersonianum* and *laterale* × *texanum* × *tigrinum* (LTTi), believed to represent separate clonal species^{8–10}, were also included. The triploid hybrids *laterale* × 2-*texanum* 2-*laterale* × *texanum*, and LTTi were chosen to include the same hybrid combinations that had been used in restriction fragment length polymorphism (RFLP) analyses^{11,12}, as well as LJJ which had not been previously examined. Some of the individual triploid hybrids chosen demonstrated genetic exchange of diagnostic electromorphs⁴. An additional species, *A. mexicanum*, was included for comparison, and the plethodontid species *Plethodon yonahlossee* was used as an outgroup. A region of 307 base pairs of the cytochrome *b* gene in each salamander was amplified and sequenced (Fig. 1).

Of the 20 hybrids, 18 form a separate monophyletic lineage

Received 6 February; accepted 12 March 1992.

1. Williams, G. C. *Sex and Evolution* (Princeton Univ. Press, Princeton, 1975).
2. Maynard Smith, J. *The Evolution of Sex* (Cambridge Univ. Press, Cambridge and New York, 1978).
3. Bell, G. *The Masterpiece of Nature* (Univ. California Press, Berkeley, 1982).
4. Densmore, L. D., Wright, J. W. & Brown, W. M. *Evolution* **43**, 943–957 (1989).
5. Moritz, C. C., Wright, J. W. & Brown, W. M. *Evolution* **43**, 958–968 (1989).
6. Densmore, L. D., Moritz, C. C., Wright, J. W. & Brown, W. M. *Evolution* **43**, 969–983 (1989).
7. Wright, J. W., Spolsky, C. & Brown, W. M. *Herpetologica* **39**, 410–416 (1983).
8. Moritz, C. et al. *Bull. N.Y. St. Mus.* **466**, 87–112 (1989).
9. Moritz, C. et al. *Genetica* (in the press).
10. Avise, J. C. & Vrijenhoek, R. C. *Molec. Biol. Evol.* **4**, 514–525 (1987).
11. Quattro, J. M., Avise, J. C. & Vrijenhoek, R. C. *Proc. natn. Acad. Sci. U.S.A.* **89**, 348–352 (1992).
12. Echelle, A. A., Dowling, T. E., Moritz, C. C. & Brown, W. M. *Evolution* **43**, 984–993 (1989).
13. Goddard, K. A., Dawley, R. M. & Dowling, T. E. *Bull. N.Y. St. Mus.* **466**, 268–297 (1989).
14. Spolsky, C. & Uzzell, T. *Molec. Biol. Evol.* **3**, 44–56 (1986).
15. Quattro, J. M., Avise, J. C. & Vrijenhoek, R. C. *Genetics* **127**, 391–398 (1991).
16. Uzzell, T. M. *Jr Copeia* **1964**, 247–300 (1964).
17. Clanton, W. *Occ. Pap. Mus. Zool. Univ. Mich.* **290**, 1–14 (1934).
18. Minton, S. A. *Herpetologica* **10**, 173–179 (1954).
19. Uzzell, T. M. Jr & Goldblatt, S. M. *Evolution* **21**, 345–354 (1967).
20. Macgregor, H. C. & Uzzell, T. M. *Jr Science* **143**, 1043–1045 (1964).
21. Bogart, J. P. *Bull. N.Y. St. Mus.* **466**, 170–179 (1989).
22. Lowcock, L. A. *Bull. N.Y. St. Mus.* **466**, 180–208 (1989).
23. Kraus, F. & Miyamoto, M. M. *Proc. natn. Acad. Sci. U.S.A.* **87**, 2235–2238 (1990).

FIG. 1 Mitochondrial DNA sequence variation from 307 base pairs of the cytochrome *b* gene in 46 salamanders, corresponding to sites 14,842–15,148 in the human sequence¹³.

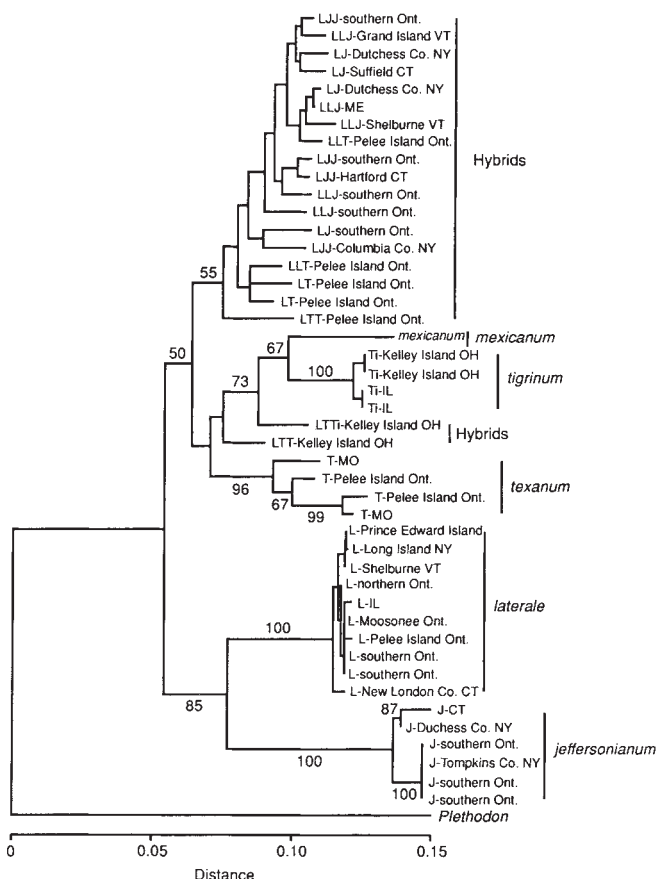
Only the 118 variable sites are shown, each identified (top) by the 3 digits that give its sequence location. Dots represent sequence identity with *Plethodon* (outgroup); N denotes ambiguities. Species are denoted Ti, *tigrinum*; T, *texanum*; L, *laterale* and J, *jeffersonianum*. Combinations of two or three abbreviations represent diploid or triploid hybrids, respectively. Sequences have been deposited in the EMBL database, accession number X63557. Specimens were from a wide range of localities (see Fig. 2).

METHODS. DNA was extracted from the liver homogenate samples (-70°C) used in allozyme analyses. Extraction, amplification (polymerase chain reaction) and sequencing followed standard protocols^{19,20}

using redesigned primers: 5'-CCAACCCCATCAACATTTTCATTATGAAA-3' and 5'-ACTGTAGCCCCTCAAAAAGATATTTGTCTCA-3'. Sequence data were read from autoradiograms and aligned using the multisequence editing program ESEE²¹.

[illegible]

1	Plethodon	TTTCCCTTGCAGCCCTCA	TAACATACATTACTCTATAAATCCACACAGTATGATACCACGATCTATAAACCAGAAATCTTACTTACTCCATCGAACCTTTAAACTAACAAATCTCA
2	L-Nort-On	C...T...T...AT...CA...TA...CA...TT...CCACACA...TGT...TTGT...C...CTA...GATGATT...TC...TCCY...C...ACT...CA...T...AG...T...NT...TACATTGTA...TG...CTCT	
3	L-PrinEdId	C...T...T...AT...CA...TA...CA...TT...CCACACA...TGT...TTGT...C...CTA...GATGATT...TC...TCCY...C...ACT...CA...TC...AG...T...T...TACATTGTA...TG...CTCT	
4	L-PeId-On	C...T...T...AT...CA...TA...CA...TT...CCACACA...G...TGT...TTGT...C...CTA...GATGATT...TC...TCCY...C...ACT...CA...T...AG...T...T...TAAATTGTA...TG...CTCT	
5	L-Sou1-On	C...T...T...AT...CA...TA...CA...TT...CCACACA...TGT...TTGT...C...CTA...GATGATT...TC...TCCY...C...ACT...CA...T...AG...T...T...TAAATTGTA...TG...CTCT	
6	L-Sou2-On	C...T...T...AT...CA...TA...CA...TT...CCACACA...TGT...TTGT...C...CTA...GATGATT...TC...TCCY...C...ACT...CA...T...AG...T...T...TAAATTGTA...TG...CTCT	
7	L-Moos-On	C...T...T...AT...CA...TA...CA...TT...CCACACA...TGT...TTGT...C...CTA...GATGATT...TC...TCCY...C...ACT...CA...T...AG...T...T...TACATTGTA...TG...CTCT	
8	L-Sheb-VI	C...T...T...AT...CA...TA...CA...TT...CCACACA...TGT...TTGT...C...CTA...GATGATT...TC...TCCY...C...ACT...CA...T...AG...T...T...TACATTGTA...TG...CTCT	
9	L-NewL-NY	C...T...T...AT...CA...TA...CA...TT...CCACACA...TGT...TTGT...C...CTA...GATGATT...TC...TCCY...C...ACT...CA...TC...AG...T...T...TACATTGTA...TG...CTCT	
10	L-LongL-NY	C...T...T...AT...CA...TA...CA...TT...CCACACA...TGT...TTGT...C...CTA...GATGATT...TC...TCCY...C...ACT...CA...TC...AG...T...T...TACATTGTA...TG...CTCT	
11	L-Illinois	C...T...T...AT...CT...TA...CT...CA...TA...TT...CCACACA...TGT...TTGT...C...CTA...GATGATT...TC...TCCY...C...ACT...CA...T...AG...T...T...TACATTGTA...TG...CTCT	
12	J-Sou1-On	C...CT...TC...A...CGGTGA...CA...CCCTTCCACACA...C...C...C...CTA...GA...TATT...T...CTT...C...ACT...AC...T...GAG...TTTACATTGTAG...TG...TCG	
13	J-Sou2-On	C...CT...TC...A...CGGTGA...CA...CCCTTCCACACA...C...C...C...CTA...GA...TATT...T...CTT...C...ACT...AC...T...GAG...TTTACATTGTAG...TG...TCG	
14	J-Sou3-On	C...CT...TC...A...CGGTGA...CA...CCCTTCCACACA...C...C...C...CTA...GA...TATT...T...CTT...C...ACT...AC...T...GAG...TTTACATTGTAG...TG...TCG	
15	J-Connecti	C...CT...TC...A...CGTGA...CA...CCCTTCCACACA...C...C...C...CTAC...GA...TATT...T...CTT...CC...ACT...G...AC...T...AG...TC...TTTACATTGTAG...TG...TCG	
16	J-Dutch-NY	C...CT...TC...A...CGTGA...CA...CCCTTCCACACA...C...C...C...CTA...GA...TATT...T...CTT...C...ACT...G...AC...T...AG...T...TTTACATTGTAG...TG...TCG	
17	J-Tompk-NY	C...CT...TC...A...CGGTGA...CA...CCCTTCCACACA...C...C...C...CTA...GA...TATT...T...CTT...C...ACT...AC...T...GAG...TTTACATTGTAG...TG...TCG	
18	T-PeIl-On	C...C...A...TA...TA...CGA...TT...CCACACA...T...T...TTTCTCA...CGA...CTATC...T...CT...CTT...C...GTC...A...T...A...TACATTGTA...TG...GCT	
19	T-Pe12-On	CCCT...C...AT...TA...TAGGCA...TT...CCACACA...T...T...TTTCTCA...CGA...CTATC...T...CT...CTT...C...ACTC...A...T...A...CT...TAGATTGTA...TG...GCT	
20	T-Missour1	C...C...A...CG...TA...CGAG...TT...CCACACA...T...T...TTTCTCA...CGATCTATC...T...CT...CTT...C...GTC...A...T...A...CT...TACATTGTA...TG...GCT	
21	T-Missour2	CCCT...C...AT...TA...TAGGCA...TT...CCACACA...T...T...TTTCTCA...CGATCTATC...T...CT...CTT...C...ACTC...A...T...A...CT...TACATTGTA...TG...GCT	
22	mexicanum	C...A...TC...AT...TA...TA...CA...TT...CCACAC...C...CCT...T...TTTCTCA...GA...CTATC...TCCT...CTT...T...A...TCC...A...GTC...A...CT...TTA...ATCGT...TGT...TCT	
23	Ti-Kel1-OH	C...A...TC...AT...TA...TA...CAG...TT...CCACACA...T...T...T...T...CTA...T...GATCTATC...T...CTT...C...A...TCC...AC...T...A...CT...TTACATCGT...TG...G.TCT	
24	Ti-Kel2-OH	C...A...TC...AT...TA...TA...CAG...TT...CCACACA...T...T...T...T...CTA...T...GATCTATC...T...CTT...C...A...TCC...AC...T...A...CT...TTACATCGT...TG...G.TCT	
25	Ti-Illino1	C...A...TC...AT...TA...TA...CAG...TT...CCACACA...T...T...T...T...CTA...T...GATCTATC...T...CTT...C...A...TCC...AC...T...A...CT...TTACATCGT...TG...G.TCT	
26	Ti-Illino2	C...A...TC...AT...TA...TA...CAG...TT...CCACACA...T...T...T...T...CTA...T...GATCTATC...T...CTT...C...A...TCC...AC...T...A...CT...TTACATCGT...TG...G.TCT	
27	LJ-Sout-On	C...A...A...CA...TA...CAG...TT...CCACACA...T...T...T...C...CTA...GA...ATC...T...CTT...T...TACTCC...A...T...A...T...G...ACACCGTA...TTG...TCT	
28	LJ-Suff-CT	C...A...A...CA...TA...CAG...CT...CCACACAC...T...T...T...C...CCA...GA...CATC...T...CTT...T...TACTCC...A...TC...A...T...ACACTGTA...TTGT...TCT	
29	LJ-Dut1-NY	C...A...A...TA...TA...CAG...CT...CCACACAC...T...T...T...C...CCA...GA...CATC...T...CTT...T...TACTCC...A...TC...A...T...A...ACTGTA...CTGT...TCT	
30	LJ-Dut2-NY	C...A...GA...CA...TA...CAG...CT...CCACACAC...T...T...T...C...CCA...GA...CATC...T...CTT...T...TACTCC...A...TC...A...T...A...ACACGTA...CTGT...TCT	
31	LT-Pe11-On	C...A...TC...A...CA...TA...CAG...CT...CCACACA...T...T...T...C...CTA...GA...CATC...T...CTT...T...TACTCC...A...T...A...CT...ACATTGTA...TTG...TCT	
32	LT-Pe12-On	C...A...CA...CA...TA...CAG...TT...CCACACA...T...T...T...C...CCA...GA...CATC...T...CTT...T...TACTCC...A...T...A...CT...TAGATTGTA...TTG...TCT	
33	LT-Pe13-On	C...A...A...CA...TA...CAG...CT...CCACACA...T...T...T...C...CTA...GATC...ATC...T...CTT...T...TGCTCC...A...T...A...CA...TACATTGTA...TTG...TCT	
34	LJ-S1-On	C...G...CA...CA...TA...CAG...CT...CCACACA...T...T...T...C...CTA...GA...ATC...T...CTT...T...TACTCC...A...T...A...CT...ACACTGTA...CTGT...TCT	
35	LJ-S2-On	C...C...CA...CA...TA...CAG...CT...CCACACA...T...T...T...C...CCA...GA...CATC...T...CTT...T...TACTCC...A...T...A...T...G...ACACTGTA...CTGT...TCT	
36	LJ-S3-On	C...C...CA...CA...TA...CAG...CT...CCACACAC...T...T...T...C...CCA...GA...CATC...T...CTT...T...TACTCC...A...TC...A...T...ACACTGTA...CTGT...TCT	
37	LJ-Har-CT	C...C...CA...CA...TA...CAG...CT...CCACACAC...T...T...T...C...CCA...GA...CATC...T...CTT...T...TACTCC...A...T...A...T...G...ACACTGTA...CTGT...TCT	
38	LJ-Col-NY	C...GC...A...CA...TA...CAG...CT...CCACACA...T...T...T...C...CCA...GA...ATC...T...CTT...T...TACTCC...A...T...A...NACACTGTA...TTG...TCT	
39	LJL-So-On	C...A...TA...TA...CAG...CT...CCACACAC...T...T...T...C...CCA...GA...CATC...T...CTT...T...TACTCC...A...TC...A...T...G...ACACTGTA...TTGT...TCT	
40	LJL-Maine	C...A...TA...TA...CAG...CT...CCACACA...T...T...T...C...CCA...GA...CATC...T...CTT...T...TACTCC...A...TC...A...T...A...ACTGTA...CTGT...TCT	
41	LJL-Gld-VT	C...C...CA...CA...TA...CAG...CT...CCACACAC...T...T...T...C...CCA...GA...CATC...T...CTT...T...TACTCC...A...TC...A...T...A...ACTGTA...CTGT...TCT	
42	LJL-She-VT	C...A...TA...TA...CAG...CT...CCACACA...T...T...T...C...CCA...GA...CATC...T...CTT...T...TACTCC...A...TC...A...T...A...ACTGTA...TTGT...TCT	
43	LT-L1-Pint	C...A...TA...TA...CAG...CT...CCACACA...CT...T...T...T...C...CCA...GA...CATC...T...CTT...T...TACTCC...A...TC...A...T...A...ACTGTA...CTGT...TCT	
44	LTT-P2-On	C...A...CNA...TAN...CA...CT...CCACACA...T...T...T...T...C...CCA...GA...CTATC...T...AT...CTT...T...TACTCCAA...TCAA...CT...TAGATTGTA...TTGN...TCT	
45	LTT-K1-On	C...A...CTA...TA...CA...CT...CCACACA...T...T...TGT...CTA...GA...CTATC...T...CTT...C...ACTC...A...T...A...CT...TAGATCGT...TG...G.TCT	
46</			



in the phylogenetic tree independent of ploidy, nuclear genomic composition, or locality (Fig. 2). The maternally derived mt DNA found in the hybrids is not from any of the four contemporary species but seems to derive from a distinct ancestral lineage. None of the 20 hybrids clustered with any of the 4 bisexual species. Analysed as characters, shared-derived nucleotides at six sites are present in all or a majority of the hybrids whereas no shared-derived sites join the hybrid lineage with *texanum* as was previously suggested^{11,12}. The four *texanum* individuals from two widely separated localities cluster as a monophyletic group at a high bootstrap *P* value (96%; Fig. 2). As expected, most of the DNA sequence differences are transitions at synonymous sites, although one site defining the major hybrid lineage is a transversion involving an amino-acid replacement (T to A resulting in methionine replacing leucine in the protein at a site corresponding to 15,050 in the human DNA sequence¹³). The two 'outlying' hybrids from Kelley Island may

FIG. 2 Evolutionary tree of the unisexual hybrid *Ambystoma* salamander complex based on mtDNA sequence data. The neighbour-joining method^{22,23} was used with the Jukes–Cantor distance²⁴ (scale bar); bootstrap *P*-values²⁵ are indicated on the tree; abbreviations are as in Fig. 1. Maximum parsimony and UPGMA analyses resulted in the same groups as indicated by vertical bars and in the same relationships, except that in the parsimony tree *texanum* was a sister group to *tigrinum* + *mexicanum* + hybrids, and in the UPGMA tree the two Kelly Island hybrids clustered and formed a sister group to *tigrinum*.

METHODS. Location of specimens: *laterale*, James Bay, Ontario Moosonee, southern Ontario (Ont.) northern Illinois (IL) Pelee Island in Lake Erie, Prince Edward Island, Vermont (VT), Connecticut (CT) and New York (NY); *jeffersonianum*, Ontario, Connecticut and New York; *texanum*, Missouri (MO) and Pelee Island, Ontario; and *tigrinum*, Kelley Island, Ohio (OH) and from northern Illinois. The programs NJTREE (L. Jin), TDRAW (W. Ferguson, University of Texas) and NJBOOT (T. S. Whittam) were used to construct the neighbour-joining tree and perform the bootstrap analysis (2,000 replications). The program UPGMA (J. C. Stephens) was used to construct a tree and to estimate the time of origin (± 2 standard errors) of the major hybrid lineage. The program PAUP (D. Swofford) was used to construct a maximum parsimony tree.

provide evidence for additional hybridization events but the data are insufficient to resolve branching order in that part of the tree.

These results are unexpected and differ from the allozyme results used to document the nuclear genomic content of the hybrids. If the female parent of a hybrid had been any one of the four species examined, the hybrid would be expected to cluster with that species. Our results confirm the uncoupling of mitochondrial and nuclear genomic evolution in this hybrid complex that was suggested by RFLP analyses of mitochondrial DNA (mtDNA)^{11,12}. However, we find the hybrids to be an ancient lineage (or lineages) that is not closely related to *texanum* as was suggested by these RFLP studies (an ancient origin for some unisexual *Ambystoma* has also been suggested by C. Spolsky, personal communication). The allozyme data show there are one or more *laterale* haploid chromosome complements in every hybrid but none of the hybrid mtDNA sequences clusters with *laterale* mtDNA sequences. Consistent with these findings, the hybrid mtDNAs exhibit considerable differentiation, as much or more than that observed within each of the bisexual species (Fig. 2). We attribute these differences from earlier results to the inclusion in this study of all four bisexual species involved in the hybrid complex, a broader geographic sampling of hybrids and bisexuals and greater resolution of genetic variation by the sequence data.

Our results answer two important, lingering questions concerning the evolution of this intriguing salamander complex: (1) there are no hybridizations where both parents are among the four bisexual species because none of the hybrids clusters with any of these species; and (2) individuals of the pure species are probably not reconstituted from the hybrids^{7,14}, because none appears within the hybrid cluster. This large difference in the evolution of mtDNA and nuclear DNA has not been reported in other unisexual vertebrates¹⁵ or in any other organism. We hypothesize that the mtDNA of the hybrids has been transmitted clonally from a very distant ancestor while the hybrids continue to acquire nuclear material from the four bisexual species.

To estimate the time of origin of the main unisexual hybrid lineage (Fig. 2), we used the calibration for cytochrome *b* sequence divergence (2.5% per million years) applied in other vertebrates^{2,16} and obtained a date of 3.9 ± 0.6 million years ago (Pliocene; independent calibration within *Ambystoma* is not possible as the fossil record is poor). The widely separated samples of *laterale* have nearly identical sequences, implying a recent origin (<200,000 years ago) which agrees with molecular data for other northern salamanders that have undergone rapid, postglacial range expansion¹⁷. The hybrids do not show such high sequence similarity; their sequence divergence suggests that multiple sublineages have existed for several million years. In contrast, the oldest unisexual vertebrate lineage previously reported, a population of Mexican poeciliid fish, is believed to be only 60,000–150,000 years old¹⁸.

Hybrid *Ambystoma* represent only one of a number of unisexual vertebrate complexes¹ and it will be interesting to see if other unisexuals have independently evolving nuclear and mitochondrial genomes. It is instructive to know that the mtDNA need not co-evolve with the nuclear genome and that phylogenies based on mtDNA may be very different from those derived from nuclear genes. □

12. Kraus, F. & Miyamoto, M. M. *Proc. natn. Acad. Sci. U.S.A.* **87**, 2235–2238 (1990).
13. Anderson, S. et al. *Nature* **290**, 457–465 (1981).
14. Bogart, J. P. in *Polyploidy: Biological Relevance* (ed. Lewis, W. H.) 341–378 (Plenum, New York, 1980).
15. Avise, J. C., Trexler, J. C., Travis, J. & Nelson, W. S. *Evolution* **45**, 1530–1533 (1991).
16. Irwin, D. M., Kocher, T. D. & Wilson, A. C. *J. molec. Evol.* **32**, 128–144 (1991).
17. Highton, R. & Webster, T. P. *Evolution* **30**, 33–45 (1976).
18. Quattro, J. M., Avise, J. C. & Vrijenhoek, R. C. *Proc. natn. Acad. Sci. U.S.A.* **89**, 348–352 (1992).
19. Kocher, T. D. et al. *Proc. natn. Acad. Sci. U.S.A.* **86**, 6196–6200 (1989).
20. Hedges, S. B., Bezy, R. L. & Maxson, L. R. *Molec. Biol. Evol.* **8**, 767–780 (1991).
21. Cabot, E. L. & Beckenbach, A. T. *Comput. appl. Biosci.* **5**, 233–234 (1989).
22. Saitou, N. & Nei, M. *Molec. Biol. Evol.* **4**, 406–425 (1987).
23. Studier, J. A. & Keppler, K. J. *Molec. Biol. Evol.* **5**, 729–731 (1988).
24. Jukes, T. H. & Cantor, C. R. in *Mammalian Protein Metabolism* (ed. Munroe, H. N.) 21–132 (Academic, New York, 1969).
25. Felsenstein, J. *Evolution* **39**, 783–791 (1985).

ACKNOWLEDGEMENTS. We thank colleagues and former students of J.P.B. for many of the specimens; M. W. Klemens and L. E. Licht for specimens collected from distant localities; T. Ota, S. Kumar and K. Maxson for help with sequence analysis; L. Tu for laboratory assistance; M. Nei for computing facilities; L. Jin, W. Ferguson, J. C. Stephens and T. S. Whittam for programs; and C. A. Hass and T. S. Whittam for comments. This work was supported by the NSERC (Canada) and NSF (USA).

Genetically engineered alteration in the chilling sensitivity of plants

N. Murata, O. Ishizaki-Nishizawa*, S. Higashi, H. Hayashi, Y. Tasaka & I. Nishida

National Institute for Basic Biology, Myodaiji, Okazaki 444, Japan

* Central Laboratories for Key Technology, Kirin Brewery Co., Shioya, Tochigi, 329-14, Japan

THE chilling sensitivity of plants is closely correlated with the degree of unsaturation of fatty acids in the phosphatidylglycerol of chloroplast membranes^{1–5}. Plants with a high proportion of *cis*-unsaturated fatty acids, such as spinach and *Arabidopsis thaliana*, are resistant to chilling, whereas species like squash with only a small proportion are not. The chloroplast enzyme glycerol-3-phosphate acyltransferase seems to be important for determining the level of phosphatidylglycerol fatty acid unsaturation^{6–9}. Here we report that the level of fatty acid unsaturation of phosphatidylglycerol and the degree of chilling sensitivity of *Nicotiana tabacum* var. Samsun (tobacco) can be manipulated by transformation with complementary DNAs for glycerol-3-phosphate acyltransferases from squash and *Arabidopsis*. The genetic manipulation of fatty acid unsaturation is known to alter the chilling sensitivity of prokaryotes¹⁰, and we have now demonstrated that it can also do so in higher plants.

Glycerol-3-phosphate acyltransferase cDNA from squash¹¹ and *Arabidopsis*¹², under the control of the cauliflower mosaic virus 35S constitutive promoter in the binary plasmid pBI-121, was introduced into tobacco plants. The cDNA encoding the full-length precursor of the enzyme from *Arabidopsis*, and containing the 5' and 3' noncoding regions¹², was inserted between the *Bam*H1 and *Sac*I sites of pBI-121, to form a plasmid designated pARA. The mature protein region of the cDNA for glycerol-3-phosphate acyltransferase from squash¹¹ was ligated with the transit region of the cDNA for the small subunit of pea Rubisco¹³. This construct was inserted between the *Bam*H1 and *Sac*I sites of pBI-121, to form a plasmid pSQ. These different constructs were introduced into *Agrobacterium tumefaciens* (LBA 4404) by electroporation and transformants of *A. tumefaciens* were selected by resistance to kanamycin and by DNA-DNA blot analysis for vector plasmids.

N. tabacum was transformed by the leaf-disk method¹⁴ and the transformed calli were selected on Murashige-Skoog (MS) medium¹⁵ containing kanamycin (100 µg ml⁻¹) and claforan (250 µg ml⁻¹). After shooting and rooting, transformants were grown at 25 °C on MS medium containing claforan (250 µg ml⁻¹) in plastic boxes under fluorescent light (50 W m⁻²) for 16 h followed by 8 h dark. Fifteen independent kanamycin-resistant plants were obtained for each construct and their phosphatidylglycerol fatty acid composition analysed. Most plants

Received 4 October 1991; accepted 29 January 1992.

1. Vrijenhoek, R. C., Dawley, R. M., Cole, C. J. & Bogart, J. P. *N.Y. State Mus. Bull.* **466**, 19–23 (1989).
2. Meyer, A., Kocher, T. D., Basasibwaki, P. & Wilson, A. C. *Nature* **347**, 550–553 (1990).
3. Bogart, J. P., Elinson, R. P. & Licht, L. E. *Science* **246**, 1032–1034 (1989).
4. Bogart, J. P. *N.Y. State Mus. Bull.* **466**, 209–217 (1989).
5. Taylor, A. S. & Bogart, J. P. *Genome* **33**, 837–844 (1990).
6. Bogart, J. P. & Licht, L. E. *Can. J. Genet. Cytol.* **28**, 605–617 (1986).
7. Lowcock, L. A. & Bogart, J. P. *Can. J. Zool.* **67**, 350–356 (1989).
8. Uzzell, T. M. & Goldblatt, S. M. *Evolution* **21**, 345–354 (1967).
9. Sessions, S. *Chromosoma* **84**, 599–621 (1982).
10. Kraus, F. *Occ. Pap. Mus. Zool. Univ. Michigan* **709**, 1–24 (1985).
11. Kraus, F. *N.Y. State Museum Bull.* **466**, 218–227 (1989).

TABLE 1 Fatty acid composition of phosphatidylglycerol from leaves of wild-type plants and transgenic tobacco plants

Plant	Fatty acid* (mol %)						$\left(\begin{smallmatrix} 18:1 \\ +18:2 \\ +18:3 \end{smallmatrix} \right)$	Cis-unsaturated molecules† (estimated)
	16:0	16:1t	18:0	18:1	18:2	18:3		
Squash	54	22	5	5	3	10	18	(36)
<i>Arabidopsis</i>	29	29	2	5	9	26	40	(80)
Tobacco	20	46	2	4	11	17	32	(64)
Transgenic tobacco with pBI-121	23	44	1	6	11	15	32	(64)
pSQ	52	29	7	2	3	7	12	(24)
pARA	24	38	2	7	12	17	36	(72)

Extraction of lipids from leaves, the separation of phosphatidylglycerol, and fatty acid analysis were as described¹. Three independent measurements were made for each plant, and the percentage deviation for (18:1 + 18:2 + 18:3) was within 2%. The same results were obtained in the other four independent plants in each group of tobacco transformants.

* Abbreviations for fatty acids: 16:0, hexadecanoic acid (palmitic acid); 16:1t, Δ^3 -trans-hexadecenoic acid; 18:0, octadecanoic acid (stearic acid); 18:1, Δ^9 -octadecenoic acid (oleic acid); 18:2, $\Delta^9,12$ -octadecadienoic acid (linoleic acid); 18:3, $\Delta^9,12,15$ -octadecatrienoic acid (α -linolenic acid).

† Cis-unsaturated molecules are phosphatidylglycerols which bind 18:1, 18:2 or 18:3 at the sn-1 position and either 16:0 or 16:1t at the sn-2 position of the backbone glycerol. The proportion of cis-unsaturated molecules can be estimated on the basis that 16:0 and 16:1t are exclusively esterified at the sn-2 position of the glycerol moiety of phosphatidylglycerol¹; the estimated values were obtained by doubling the value for the proportion of (18:1 + 18:2 + 18:3).

showed large changes in fatty acid unsaturation. Five plants from each group that showed large changes in fatty acid unsaturation were selected for further analysis.

Hybridization of genomic DNA from leaves of these transformants showed that the cDNAs were stably integrated into the genome, with the number of integrated copies estimated to be in the range of 1–5. RNA–DNA hybridization of poly(A)⁺ RNA from leaves of the transformed plants demonstrated that the glycerol-3-phosphate acyltransferase messenger RNA was highly expressed.

Using antibodies against the glycerol-3-phosphate acyltransferases from *Arabidopsis* and squash, we analysed the protein overexpression by immunoblotting¹⁶. In plants transformed with *Arabidopsis* or squash cDNAs, the protein of mature size was detected in leaves and the chloroplast-soluble fraction of transformants. The level of protein was estimated as 0.1–1% of the total leaf protein. This is much higher than the endogenous level of the native protein of glycerol-3-phosphate acyltransferase, which is <0.01% of the total soluble protein of leaves⁶. These observations demonstrate that the protein is transported into chloroplasts and is processed to the mature enzyme.

We analysed the lipids of these transformed plants and found no discernible changes in overall fatty acid composition. In the separated lipid classes, only the fatty acid composition of phosphatidylglycerol was significantly altered by transformation with glycerol-3-phosphate acyltransferase cDNA (Table 1). When the squash enzyme was introduced, the content of cis-unsaturated fatty acids (18:1 + 18:2 + 18:3; see Table 1 for nomenclature) in phosphatidylglycerol was decreased. By contrast, introduction of the *Arabidopsis* enzyme caused a small increase in these fatty acids. These results are consistent with the *in vivo* content of these fatty acids being determined solely by glycerol-3-phosphate acyltransferase. Composition of the lipid classes was not altered by introduction of the enzyme (Table 2). These results indicate that overexpression of glycerol-3-phosphate

acyltransferase in the transformed plants does not affect the flux through the two pathways of glycerolipid synthesis^{17,18}, but alters the selectivity for the saturated and cis-unsaturated fatty acids at the sn-1 position of phosphatidylglycerol.

The sensitivity to chilling of transformed plants was assayed by two methods: photoinhibition of photosynthesis at chilling temperatures, and visible damage of the whole plant at chilling temperatures. In the former, leaves were exposed to 1 °C for 4 h

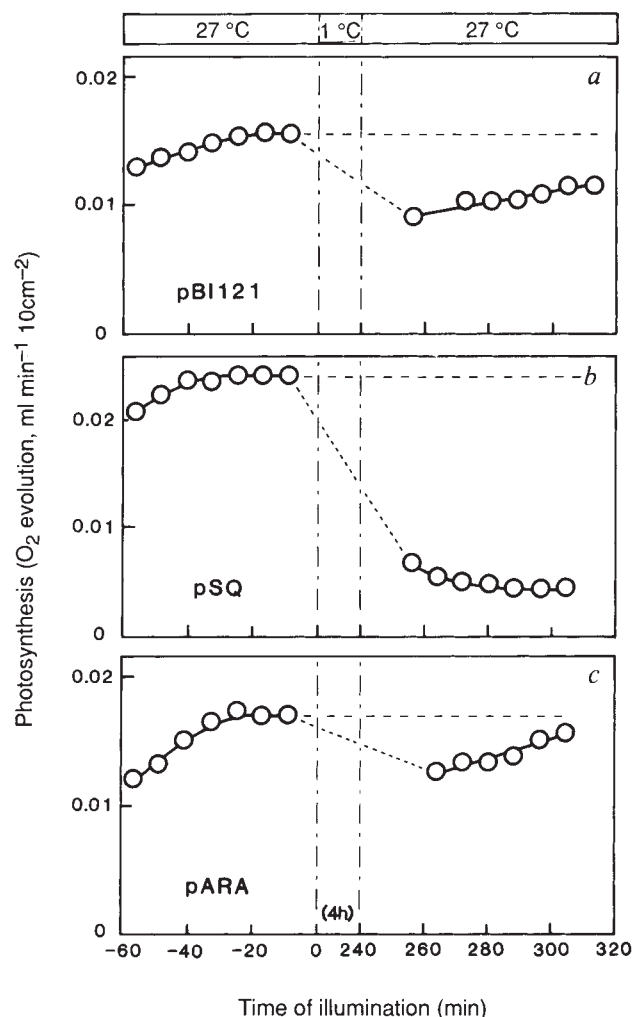


FIG. 1 Chilling-induced damage to the photosynthetic activity of leaves of transgenic tobacco plants. a, Transformant with pBI-121 (control; Clontech). b, Transformant with pARA. c, Transformant with pSQ.

METHODS. A 10-cm² disk obtained from an intact leaf, about 4 cm × 6 cm, was placed in the temperature-controlled chamber equipped with an oxygen electrode (Hansatech, LD2). The leaf disk was illuminated during the whole experiment at 27 °C and 1 °C with white light at 300 W m⁻² (sunlight intensity) from a 100-W tungsten lamp (Hansatech LS2) through a heat-absorbing optical filter (Hoya, HA-50). Oxygen-evolving activity of the leaf disk at 27 °C was measured at 8-min intervals.

TABLE 2 Composition of membrane lipids from leaves of transgenic tobacco plants

Tobacco plant	Lipid class* (mol %)						
	MGDG	DGDG	SQDG	PG	PC	PE	PI
Wild type (control)	45	27	3	7	11	5	2
Transformant with pARA	44	26	4	6	12	6	2
pSQ	43	27	4	7	12	6	1

Extraction of lipids, the separation of lipid classes, and fatty acid analysis have been described¹. The proportions of individual lipid classes were determined on the basis of fatty acid content. Two independent determinations were made for each plant, and the percentage deviation was within 2% in MGDG and DGDG, and within 1% in SQDG, PG, PC, PE and PI. The same results were obtained for the other four independent plants in each group of transformants.

* Abbreviations for lipid classes: DGDG, digalactosyldiacylglycerol; MGDG, monogalactosyldiacylglycerol; PC, Phosphatidylcholine; PE, phosphatidylethanolamine; PG, phosphatidylglycerol; PI, phosphatidylinositol; SQDG, sulphoquinovosyldiacylglycerol.

in the light before returning them to 27 °C and monitoring the effects of this treatment on photosynthetic oxygen evolution (Fig. 1). The transformant control (Fig. 1a) and wild-type tobacco plants showed some chilling sensitivity. Upon introduction of the squash enzyme, this sensitivity was increased (Fig. 1b). But, introduction of the *Arabidopsis* enzyme decreased chilling sensitivity (Fig. 1c). Four independent experiments showed that the extent of inactivation by chilling, measured by the activity 60 min after the chilling treatment relative to that before the chilling treatment, was $25 \pm 11\%$, $88 \pm 12\%$ and $7 \pm 3\%$ in transformants with pBI-121 (control), pSQ and pARA, respectively. The degree of chilling sensitivity of tobacco plants transformed with pSQ and pARA, are similar to those of squash and spinach, respectively (data not shown).

For the second chilling-sensitivity assay of visible damage to the whole plant, the different transformed plants were exposed to 1 °C for 10 days. Damage such as chlorosis and total deterioration of the leaves was obvious after the leaves were kept at 25 °C for 2 days after chilling treatments (Fig. 2). Leaves of the wild-type plant and the plant transformed with pBI-121 (control) suffered partial chlorosis by chilling treatment. The plant transformed with pSQ was severely damaged and some leaves were completely dead. In contrast, the plant transformed with pARA was much more resistant to chilling than the wild type and the plant transformed with pBI-121. These differences in the chilling sensitivity correlated with the extent of fatty acid unsaturation in phosphatidylglycerol.

These results demonstrate that it is possible to alter the satur-

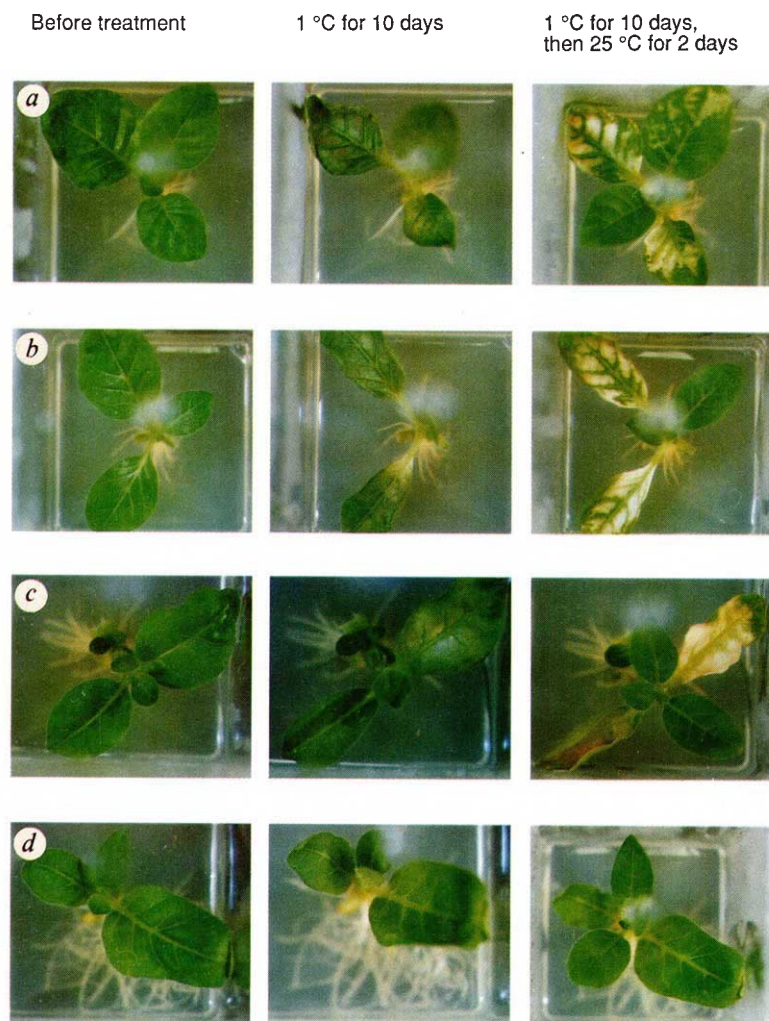


FIG. 2 Chilling-induced visible damage of transgenic tobacco plants. a, Wild type; b, Transformant with pBI-121 (control); c, Transformant with pARA; d, Transformant with pSQ.

METHODS. Plants, grown at 25 °C with 16 h of light and 8 h of dark, were exposed to a temperature of 1 °C for 10 days and returned to the same growth conditions. During exposure at 1 °C, plants were continuously illuminated by white fluorescent lamps at an intensity of 20 W m⁻².

ation of the fatty acids of phosphatidylglycerol by the introduction of an appropriate acyltransferase. They also define the *in vivo* function of acyltransferases. Factors apart from the extent of unsaturation of the fatty acids in phosphatidylglycerol may affect chilling sensitivity, but our experiments have shown that it is an important contributor. □

Received 16 January; accepted 28 February 1992.

1. Murata, N., Sato, N., Takahashi, N. & Hamazaki, Y. *Pl. Cell Physiol., Tokyo* **23**, 1071-1079 (1982).
2. Murata, N. *Pl. Cell Physiol., Tokyo* **24**, 81-86 (1983).
3. Roughan, G. P. *Pl. Physiol.* **77**, 740-746 (1985).
4. Tasaka, Y., Nishida, I., Higashi, S., Beppu, T. & Murata, N. *Pl. Cell Physiol., Tokyo* **31**, 545-550 (1990).
5. Murata, N. & Nishida, I. in *Chilling Injury of Horticultural Crops* (ed. Wang, C. Y.) 181-199 (CRC, Florida, 1990).

6. Nishida, I., Frentzen, M., Ishizaki, O. & Murata, N. *Pl. Cell Physiol., Tokyo* **28**, 1071-1079 (1987).
7. Frentzen, M., Nishida, I. & Murata, N. *Pl. Cell Physiol., Tokyo* **28**, 1195-1201 (1987).
8. Bertram, M. & Heinz, E. *Pl. Physiol.* **88**, 653-657 (1981).
9. Frentzen, M., Heinz, E., McKeon, T. A. & Stumpf, P. K. *Eur. J. Biochem.* **129**, 629-636 (1983).
10. Wada, H., Gombos, Z. & Murata, N. *Nature* **347**, 200-203 (1990).
11. Ishizaki, O., Nishida, I., Agata, K., Eguchi, G. & Murata, N. *FEBS Lett.* **238**, 424-430 (1988).
12. Nishida, I. et al. in *Plant Lipid Biochemistry, Structure and Utilization* (eds Quinn, P. J. & Harwood, J. L.) 462-464 (Portland, London, 1991).
13. Schreier, P. H., Seftor, E. A., Schell, J. & Bohnert, H. J. *EMBO J.* **4**, 25-32 (1985).
14. Horsch, R. B. et al. *Science* **223**, 496-498 (1984).
15. Murashige, J. & Skoog, F. *Physiol. Pl.* **115**, 473-497 (1962).
16. Sambrook, J., Fritsch, E. F. & Maniatis, T. *Molecular Cloning: A Laboratory Manual* 2nd edn (Cold Spring Harbor Laboratory, New York, 1989).
17. Roughan, G. P. & Slack, C. R. A. *Rev. Pl. Physiol.* **33**, 97-132 (1982).
18. Browne, J., Warwick, N., Somerville, C. R. & Slack, C. R. *Biochem. J.* **235**, 25-31 (1986).

ACKNOWLEDGEMENTS. We thank Y. Makino for measurement of oxygen evolution, and T. Slabas (University of Durham) for critically reading the manuscript.

Aberrant regulation of *ras* proteins in malignant tumour cells from type 1 neurofibromatosis patients

Tanya N. Basu*, David H. Gutmann†, Jonathan A. Fletcher‡, Thomas W. Glover§, Francis S. Collins† & Julian Downward*||

* Signal Transduction Laboratory, Imperial Cancer Research Fund, Lincoln's Inn Fields, London WC2A 3PX, UK

† Howard Hughes Medical Institute and Departments of Internal Medicine and Human Genetics, University of Michigan Medical School, Ann Arbor, Michigan 48109, USA

‡ Department of Pathology, Brigham and Women's Hospital, 75 Francis Street, Boston, Massachusetts 02115, USA

§ Departments of Pediatrics and Human Genetics, University of Michigan Medical School, Ann Arbor, Michigan 48109, USA

DEFECTS in the *NF1* gene have been implicated in the inherited disorder neurofibromatosis type 1, which is characterized by several developmental abnormalities including an increased frequency of benign and malignant tumours of neural crest origin (neurofibromas and neurofibrosarcomas respectively)¹. The *NF1* gene encodes a ubiquitous protein homologous to p120^{GAP}, the GTPase-activating protein (GAP) for the products of the *ras* proto-oncogenes²⁻⁶. When expressed in non-mammalian systems, the region of the *NF1* gene homologous to p120^{GAP} produces a protein with GAP-like activity⁷⁻⁹. Here we present evidence that the *ras* proteins in malignant tumour cell lines from patients with type 1 neurofibromatosis are in a constitutively activated state, as judged by the guanine nucleotide bound to them, and are necessary for cellular proliferation. These cells contain p21^{ras} and p120^{GAP} that are both functionally wild type, but barely any functional NF1 protein. Our results show that the NF1 protein is normally essential for correct negative regulation of *ras* proteins in the cell, even in the presence of normal p120^{GAP}, and they support the hypothesis that *NF1* is a tumour-suppressor gene whose product acts upstream of *ras*.

The activation state of p21^{ras} within a cell can be determined by measuring the ratio of GTP (active) to GDP (inactive) bound to it. Transforming mutants of p21^{ras} bind large amounts of GTP, whereas wild-type p21^{ras} is almost entirely GDP-bound unless the cell has received one of several extracellular stimuli (reviewed in ref. 10). The proportion of GTP to total guanine nucleotide bound to p21^{ras} in a number of cell lines is shown

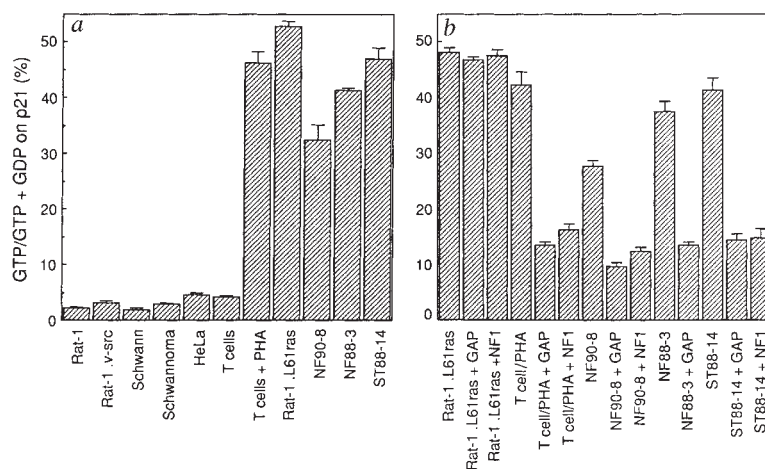
in Fig. 1a. Normal Rat-1 fibroblasts, HeLa cells, primary rat Schwann cells, schwannoma cells and quiescent human T cells all have very small amounts of GTP bound to p21^{ras} (less than 5% in the absence of serum). By contrast, Rat-1 cells transformed by an activated *ras* mutant encoding leucine at position 61 (Leu-61 *ras*) have $\geq 50\%$ of GTP-bound p21^{ras}. A similar ratio was found for p21^{ras} from phytohaemagglutinin-activated T cells¹¹. This ratio was measured for three cell lines, probably of Schwann-cell origin, derived from neurofibrosarcomas (malignant schwannomas) from neurofibromatosis type 1 patients (NF 90-8, NF 88-3 and ST 88-14): each had high levels of GTP on p21^{ras} (up to 50% in the case of ST 88-14).

The increased activation state of p21^{ras} in the neurofibrosarcoma cells could either be due to a defect in a protein normally involved in the regulation of p21^{ras} or to a mutation in a *ras* gene itself. To investigate the possibility of a *ras* mutation, p21^{ras} was immunoprecipitated from [³²P]orthophosphate-labelled cells using the monoclonal antibody Y13-259. This antibody inhibits the biological function of p21^{ras} and greatly reduces the ability of GAP to stimulate GTP hydrolysis on p21^{ras}, but it does not completely block the interaction of bound p21^{ras} with GTPase-activating proteins. We then exposed the washed immunoprecipitates to recombinant GAP or NF1 catalytic domains. The amount of GTP bound to p21^{ras} from Rat-1 cells transformed by Leu-61 *ras* was not affected by this treatment as activated mutants of p21^{ras} are insensitive to the GTPase-activating activity of GAP or NF1 (see Fig. 1b), whereas the GTP levels on the endogenous normal p21^{ras} from phytohaemagglutinin-activated peripheral blood T-lymphoblasts decreased significantly. The p21^{ras} from the neurofibrosarcoma cell lines was also sensitive to GAP and to NF1 proteins in this assay (see Fig. 1b). Therefore p21^{ras} in the neurofibrosarcoma cells behaves normally and shows no evidence of having undergone activating mutation, although some weakly activating mutations may not be evident in this assay.

If there are no functional defects in the *ras* proteins themselves, it is likely that the *ras* regulatory mechanisms are defective in neurofibrosarcoma cells. We therefore measured the enzymic activity of the GTPase-activating proteins p120^{GAP} and NF1 GAP-related protein (GRP) in lysates and immunoprecipitates from the neurofibrosarcoma cell lines, HeLa cells, NIH3T3 cells and a transformed Schwann cell line from rat (Fig. 2). Lysates from all the neurofibrosarcoma cell lines have high levels of overall GAP-like activity, albeit less than from other cell types. Immunoprecipitates of p120^{GAP} with an antiserum (PW6) raised against a non-catalytic region of the molecule also have GAP-like activity which is comparable to that in similar samples from other cell types. By contrast, immunoprecipitates of NF1 GRP using an antiserum against the carboxyl terminus of the protein (D1) do not have this GAP-like activity when they are produced from lysates of the neurofibrosarcoma cells, but are very active when derived from the other cell lines.

|| To whom correspondence should be addressed.

FIG. 1 Analysis of the guanine nucleotide bound to endogenous p21^{ras} and its sensitivity to GAP in neurofibrosarcoma cells. Cells were labelled with [³²P]orthophosphate, lysed and p21^{ras} was immunoprecipitated using anti-Ras monoclonal antibody Y13-259. **a**, Amount of GTP as a proportion of total nucleotide recovered on p21^{ras} in various cell types. The T cells were peripheral blood human T lymphoblasts which were either quiescent or treated with 10 µg ml⁻¹ phytohaemagglutinin for 10 min. **b**, p21^{ras} immunoprecipitates were exposed as indicated to 20 µg ml⁻¹ purified bacterially expressed human GAP or NF1 catalytic domain for 100 min at 30 °C before analysis of bound nucleotide. **METHODS.** Cells (5 × 10⁶) were labelled for 16 h in the absence of serum, lysed and p21^{ras} immunoprecipitated using the procedure described in ref. 11, with the following modifications. Cells were lysed in buffer containing 1% Triton X-114 instead of Triton X-100; after removal of cell debris by centrifugation, phases were partitioned in the presence of 0.5 M NaCl for 2 min at 37 °C. The detergent gel containing p21^{ras} was recovered by spinning down for 2 min at 14,000g at 20 °C and redissolved in lysis buffer containing 1% Triton X-100, 0.5% sodium deoxycholate, 0.05% sodium dodecyl sulphate and 0.5 M NaCl. Washed immunoprecipitates were incubated in 10 µl 50 mM HEPES buffer, pH 7.5, plus 5 mM MgCl₂ and 1 mg ml⁻¹ bovine serum albumin, with or without 0.2 µg GAP or NF1 protein. GAP protein was made as a fusion protein in pGEX-2T (Pharmacia) using the product of a polymerase chain reaction designed to insert specific cloning sites. The fusion protein



contained amino acids 702–1,044 of the human GAP sequence. NF1 catalytic domain was expressed in the same vector as described in ref. 7. The Rat 1-derived cell lines have been described¹⁹. Rat Schwann cells were grown as primary cultures¹⁷. T cells were prepared as in ref. 20. The rat schwannoma is line 33B. All determinations are the mean of three experiments with standard errors indicated.

Neurofibrosarcoma cells therefore seem to contain functional p120^{GAP} but no functional NF1 GRP, whereas the other cell types contain both. These data also indicate that the full-length NF1 GRP is a GTPase-activating protein for p21^{ras}; previously it has been shown that antibodies raised against the carboxy-terminal sequence of NF1 GRP can liberate GAP activity from a detergent-insoluble fraction of bovine brain¹².

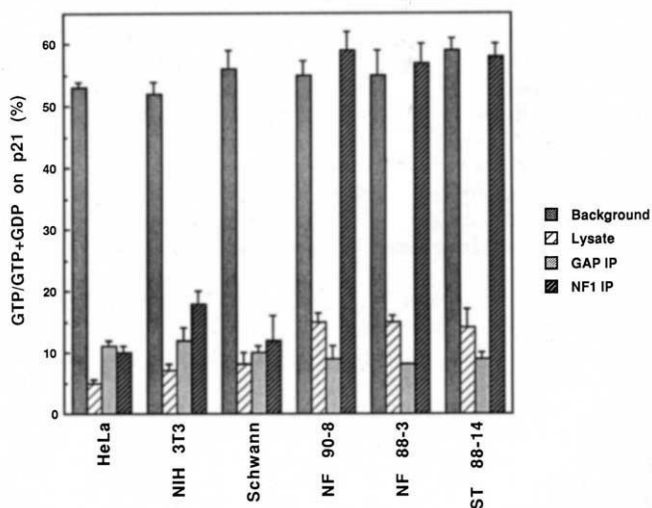
To analyse the p120^{GAP} and NF1 GRP present in these cell types, we ran immunoprecipitates similar to those used to assay GAP activity on SDS-polyacrylamide gels. In the case of p120^{GAP}, [³⁵S]methionine-labelled lysates were immunoprecipitated with antibody PW6; we found that all the neurofibrosarcoma cell lines expressed p120^{GAP} of the expected molecular weight and to the same extent as HeLa cells (Fig. 3a). For NF1 GRP, we raised a second antibody, G1, against the catalytic domain of NF1 GRP which did not crossreact with p120^{GAP}, and used it to probe immunoblots of D1 immunoprecipitates (Fig. 3b). In HeLa cells this combination of the two antisera D1 and G1 reveals an NF1 GRP with an apparent relative

molecular mass (M_r) of ~250,000 (250K)¹³. In the neurofibrosarcoma cells we could not detect this protein initially, although when five times as much ST 88-14 lysate was used and the immunoblots were overexposed, a faint band of the correct M_r could be detected (Fig. 3c). The level of expression is ~2% of that in HeLa cells. No expression of NF1 GRP could be found in NF90-8 or NF88-3 cells, nor in whole lysates of any neurofibrosarcoma cells immunoblotted directly with G1 (data not shown).

The neurofibrosarcoma cell lines used here are all derived from malignancies from patients with neurofibromatosis type 1. ST 88-14 cells are transformed, as established by their ability to form small colonies in soft agar and to grow in low serum concentrations (data not shown). ST 88-14 was derived from a tumour containing a translocation at chromosome 17 band q11.2, close to the site of *NF1*; this neoplasm contained no normal chromosome 17 (case 7 in ref. 14). There is, however, no evidence for disruption of the *NF1* coding sequence in ST 88-14 (D. Marchuk and F.C., unpublished data). NF 88-3 has undergone

FIG. 2 GAP-type activity of p120^{GAP} and NF1 GRP in immunoprecipitates (IP) from neurofibrosarcoma and control cells. The p120^{GAP}-specific antiserum PW6 and the NF1 GRP-specific antiserum D1 were used to immunoprecipitate these proteins from cell lysates. Washed immunoprecipitates were tested for their ability to stimulate the hydrolysis of [α -³²P]GTP bound to added p21^{ras}. The proportion of GTP to total guanine nucleotide bound at the end of the assay is shown.

METHODS. Cells (2 × 10⁷) were lysed in 2 ml 1% Triton X-100, 0.5% sodium deoxycholate, 50 mM Tris-HCl, pH 7.5, 100 mM NaCl, 1 mM EGTA, 10 mM benzamidine, 10 µg ml⁻¹ trypsin inhibitor, 10 µg ml⁻¹ aprotinin, 10 µg ml⁻¹ leupeptin, 1 mM dithiothreitol, 1 mg ml⁻¹ BSA. After centrifugation (for 5 min at 14,000g and 4 °C), 300 µl lysate were immunoprecipitated for 90 min at 0 °C in duplicate with 5 µg of the antibody indicated or with non-specific IgG (background). Protein A-agarose (10 µl) was then added and the mixture tumbled for 30 min at 4 °C. Immune complexes were washed four times with 1 ml 50 mM HEPES, pH 7.5, 0.5 M NaCl, 5 mM MgCl₂, 0.1% Triton X-100, 0.005% SDS, and once with 1 ml 50 mM HEPES, pH 7.5, 5 mM MgCl₂. To the immune complex was added 0.25 ng p21^{Ha-ras} bound to 1 µCi of [α -³²P]GTP. The mixture was incubated for 30 min at 30 °C and the reaction stopped by addition of 0.3 ml cold 50 mM HEPES, pH 7.5, 0.5 M NaCl, 5 mM MgCl₂, 0.1% Triton X-100, 0.005% SDS, 1 mg ml⁻¹ BSA. The p21^{ras} was immunoprecipitated from the supernatant using antibody Y13-259. Other procedures have been described¹¹. In addition, 50 µl lysate was assayed directly.



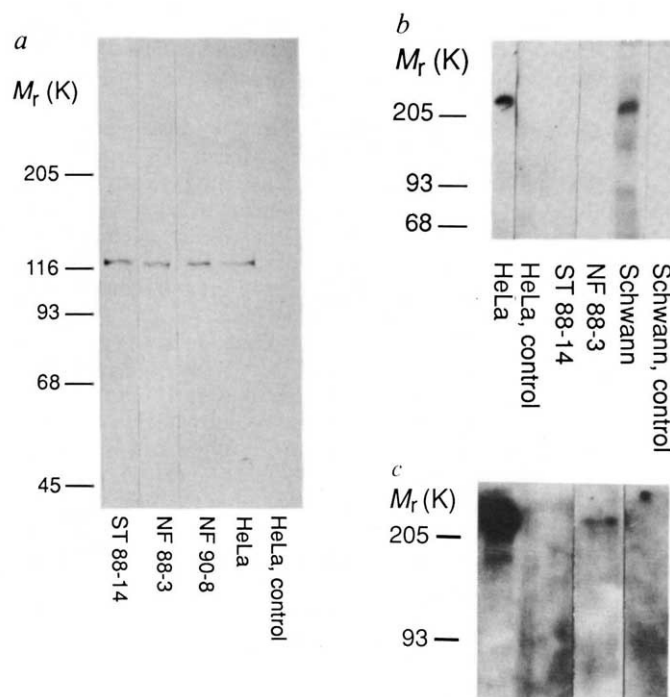


FIG. 3 Characterization by immunoprecipitation and western blotting of p120^{GAP} and NF1 GRP in neurofibrosarcoma cells. **a**, Lysates from cells labelled with [³⁵S]methionine were immunoprecipitated using antibodies against p120^{GAP} (PW6). **b**, Lysates from unlabelled cells were immunoprecipitated with antibodies against the carboxyl terminus of NF1 GRP (D1) and then immunoblotted using antibodies against the catalytic domain of NF1 GRP (G1). **c**, Longer exposure of immunoblot of NF1 GRP immunoprecipitates from HeLa cells and fivefold excess of ST 88-14 cells.

METHODS. Labelling and immunoprecipitation are described in ref. 19. Each lane in **a** and **b** represents lysates from 10⁷ cells. PW6 was an affinity-purified rabbit antiserum raised against a pGEX fusion protein containing residues 350–702 of human p120^{GAP} sequence (C. Allbright and R. Weinberg, personal communication). Antiserum D1 is described in ref. 13. Antiserum G1 was raised against a peptide from the GAP catalytic domain of NF1 GRP (residues 1,070–1,087, using the numbering in ref. 5) and immunoblotting was as described in ref. 21, using [¹²⁵I]-labelled donkey antirabbit IgG. Lanes marked 'control' were immunoprecipitated with nonspecific rabbit IgG (**a**) or preimmune D1 IgG (**b**, **c**). In **c**, each track represents 10⁷ HeLa cells or 5 × 10⁷ ST 88-14 cells; exposure time was five times longer than in **b**. Migration of M_r standards is shown on the left of each blot.

allele loss on 17q (ref. 15) and has a mutation in exon 4 of the p53 gene¹⁶. The defects in the *NF1* genes of these cells are unknown.

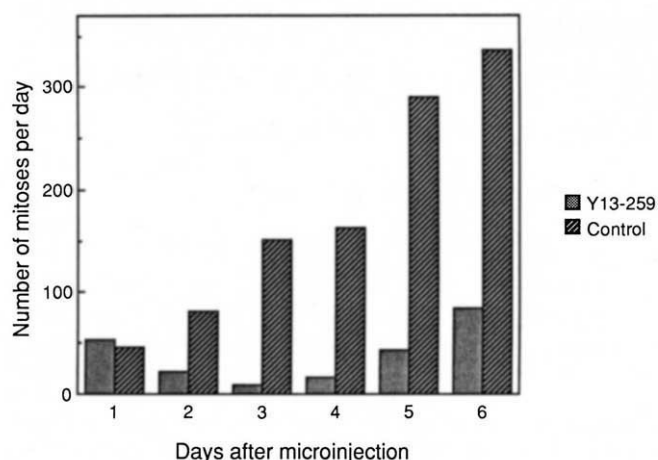


FIG. 4 Inhibition of neurofibrosarcoma cell growth by microinjection of antibodies against p21^{ras}. ST 88-14 cells grown in the presence of 15% fetal calf serum were microinjected with either 1 mg ml⁻¹ Y13-259 anti-p21^{ras} monoclonal antibody or 1 mg ml⁻¹ control immunoglobulin²². Mitotic events in a field of ~50 injected cells were followed using time-lapse video recording. The number of mitoses in each 24-h period is plotted as a proportion of the number of cells injected.

It has been proposed that p21^{ras} acts upstream of NF1 GRP because *ras* oncogenes can act alone to inhibit proliferation of Schwann cells, although in combination with nuclear oncogenes they are transforming¹⁷. Because there are frequent mutations in p53 in neurofibrosarcomas from NF1 patients¹⁸, including at least one of the three studied here, a model in which NF1 GRP is a downstream target of a growth-inhibitory p21^{ras} cannot explain the malignancies that are features of NF1. Furthermore, microinjection of the p21^{ras}-neutralizing monoclonal antibody Y13-259 strongly inhibits growth of ST 88-14 cells (Fig. 4); p21^{ras} therefore has a positive effect on proliferation in this system and the absent NF1 GRP cannot be an essential downstream effector (barring the unlikely possibility that the extremely low residual levels of NF1 GRP are sufficient to carry such a signal).

Our data show that these malignant neurofibrosarcoma cells do not produce functional full-length NF1 GRP, or only very low levels of it. The loss of NF1 GRP causes a constitutive activation of p21^{ras}, even in the presence of normal levels of fully functional p120^{GAP}. This suggests that NF1 GRP is the primary negative regulator of p21^{ras}, at least in neural crest-derived cells, possibly owing to the higher affinity of p21^{ras} for NF1 GRP than for p120^{GAP} (ref. 8), and to our knowledge is the first indication that NF1 GRP is involved in the negative control of *ras* proteins in whole cells and hence in the inhibition of cell proliferation. □

Received 10 January; accepted 6 March 1992.

- Riccardi, V. & Eichner, J. *Neurofibromatosis: Phenotype, Natural History and Pathogenesis* (Johns Hopkins Univ. Press, Baltimore, 1986).
- Viskochil, D. et al. *Cell* **62**, 187–192 (1990).
- Cawthorn, R. M. et al. *Cell* **62**, 193–201 (1990).
- Wallace, M. R. et al. *Science* **249**, 181–186 (1990).
- Xu, G. et al. *Cell* **62**, 599–608 (1990).
- Buchberg, A. M., Cleveland, L. S., Jenkins, N. A. & Copeland, N. G. *Nature* **347**, 291–294 (1990).
- Xu, G. et al. *Cell* **63**, 835–841 (1990).
- Martin, G. A. et al. *Cell* **63**, 843–849 (1990).
- Ballester, R. et al. *Cell* **63**, 851–859 (1990).
- Downward, J. *Curr. Opin. dev. Genet.* **2**, 13–18 (1992).
- Downward, J., Graves, J. D., Warne, P. H., Rayter, S. & Cantrell, D. A. *Nature* **346**, 719–723 (1990).
- Hattori, S. et al. *Biochem. biophys. Res. Commun.* **177**, 83–89 (1991).
- Gutmann, D. H., Wood, D. L. & Collins, F. S. *Proc. natn. Acad. Sci. U.S.A.* **88**, 9658–9662 (1991).

- Fletcher, J. A. et al. *New Engl. J. Med.* **324**, 436–443 (1991).
- Glover, T. W. et al. *Genes, Chromosomes Cancer* **3**, 62–70 (1991).
- Nigro, J. M. et al. *Nature* **342**, 705–708 (1989).
- Ridley, A. J., Paterson, H. F., Noble, M. & Land, H. *EMBO J.* **7**, 1635–1645 (1988).
- Menon, A. G. et al. *Proc. natn. Acad. Sci. U.S.A.* **87**, 5435–5439 (1990).
- Downward, J., de Gunzburg, J., Riehl, R. & Weinberg, R. A. *Proc. natn. Acad. Sci. U.S.A.* **85**, 5774–5778 (1988).
- Cantrell, D. A., Collins, M. K. & Crumpton, M. J. *Immunology* **65**, 343–349 (1988).
- Harlow, E. & Lane, D. *Antibodies: A Laboratory Manual* (Cold Spring Harbor Laboratory, New York, 1988).
- Smith, M. R., DeGudicibus, S. J. & Stacey, D. W. *Nature* **320**, 540–543 (1986).

ACKNOWLEDGEMENTS. We thank H. Land, C. Barth and S. Dawkins for help with Schwann cells, C. Gilbert for microinjection and F. Tamanoi for the NF1 expression vector. This work was partly supported by a grant from the NIH. F.S.C. is an investigator and D.H.G. is an associate of the Howard Hughes Medical Institute.

Spatial structure of cone inputs to receptive fields in primate lateral geniculate nucleus

R. Clay Reid* & Robert M. Shapley

New York University, Center for Neural Science, New York, New York 10003, USA

HUMAN colour vision depends on three classes of cone photoreceptors, those sensitive to short (S), medium (M) or long (L) wavelengths, and on how signals from these cones are combined by neurons in the retina and brain. Macaque monkey colour vision is similar to human, and the receptive fields of macaque visual neurons have been used as an animal model of human colour processing¹. P retinal ganglion cells and parvocellular neurons are colour-selective neurons in macaque retina and lateral geniculate nucleus. Interactions between cone signals feeding into these neurons are still unclear. On the basis of experimental results with chromatic adaptation, excitatory and inhibitory inputs from L and M cones onto P cells (and parvocellular neurons) were thought to be quite specific^{2,3} (Fig. 1a). But these experiments with spatially diffuse adaptation did not rule out the 'mixed-surround' hypothesis: that there might be one cone-specific mechanism, the receptive field centre, and a surround mechanism connected to all cone types indiscriminately (Fig. 1e). Recent work has tended to support the mixed-surround hypothesis⁴⁻⁸. We report here the development of new stimuli to measure spatial maps of the linear L-, M- and S-cone inputs to test the hypothesis definitively. Our measurements contradict the mixed-surround hypothesis and imply cone specificity in both centre and surround.

Receptive fields were mapped with stimuli made up of arrays of square regions (pixels) whose luminance or colour was modulated in a pseudo-random fashion⁹. Each frame of the stimulus looks like a scrambled chess board in which each square is one of two colours (or shades of grey) which seem to be assigned at random. The rapidly modulated stimulus resembles a coarse version of television 'noise'. We correlated nerve impulses with the patterns that preceded them¹⁰⁻¹⁴. For a given time interval between stimulus and nerve impulse, the linear receptive field was defined as the spatial stimulus that on average preceded each impulse.

Colour versions of the pseudo-random stimulus were used to map the responses from each of the cone types with the method of silent substitution¹⁵⁻¹⁷. With three different primary colours (red, green and blue phosphors of the cathode ray tube), stimulus colour pairs can be constructed such that the transition between them is visible to the photopigment in one cone type but is invisible (silent) to the other two cone photopigments.

Results were obtained from 28 red-green parvocellular neurons and 8 magnocellular neurons in the lateral geniculate nucleus (LGN) of five monkeys. For 10 of these neurons we recorded simultaneously from synaptic S-potentials which correspond to nerve impulses in retinal ganglion cell axons, the inputs to LGN neurons^{18,19}. For each S-potential/LGN spike pair, the receptive field for the LGN neuron and its retinal input had the same qualitative structure.

The main results on parvocellular receptive fields are illustrated in Fig. 2 for a typical colour-opponent (green-off/red-on type I; ref. 2) parvocellular neuron. The receptive field in response to luminance modulation (Fig. 2a) had a centre-surround organization. It was derived by calculating the average achromatic spatial stimulus that appeared to this neuron 59 ms before each action potential. The off-centre was due to M cones (Fig. 2c), whereas the on-surround was driven only by L cones

(Fig. 2b). The size and the signature of the surround are best appreciated by noting the region where the receptive field is lighter than the background grid (that is, where the grid appears dark in a and b).

Cone-isolating receptive fields in Fig. 2 are also represented as the average colour stimulus that preceded each response (Fig. 2d, e). The parvocellular neuron in Fig. 2 always responded to reddening, whether it was a decrement for M cones (e), or an increment for L cones (d); these figures also illustrate the colours of cone-isolating stimuli. Each pixel of the cone-isolating stimuli alternated between a red and green (upper left and right corners) that were detectable by only one cone class.

One striking feature of L- and M-cone inputs in Fig. 2 is similarity of spatial organization. Of 28 red-green parvocellular neurons studied, 11 had spatially coextensive L- and M-cone inputs and non-opponent luminance receptive fields (type II; ref. 2). The remaining 17 cells, like the one in Fig. 2, had spatially opponent luminance receptive fields (type I; ref. 2). Even for these neurons, most of the antagonistic input from the surround cone class was coextensive with the centre. Note that the M-cone receptive field in Fig. 2c and e shows a centre, but no antagonistic surround. This result was seen for almost all red-green parvocellular neurons and strongly supports the cone-specific surround over the mixed surround model (compare Fig. 1c and g).

We devised a numerical measure, the ratio of surround to centre strength, to make our results quantitative (Fig. 3). This ratio would be negative for receptive fields with a centre-surround organization. It would be positive or near zero for non-opponent receptive fields. For luminance stimuli, the ratio was often negative for both parvocellular and magnocellular neurons. But parvocellular neurons had ratios that were positive or near zero for almost all cone-isolating stimuli. For our parafoveal and near peripheral sample of parvocellular receptive fields, L-cone input produced a significant negative surround-centre ratio for only 2 of 28 neurons; the ratio for M-cone inputs was always non-negative, indicating no spatial opponency. Therefore, our data do not support the random connection or

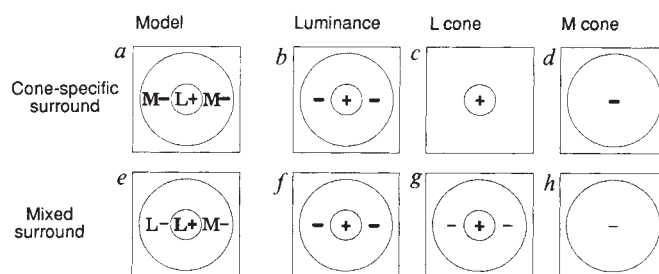


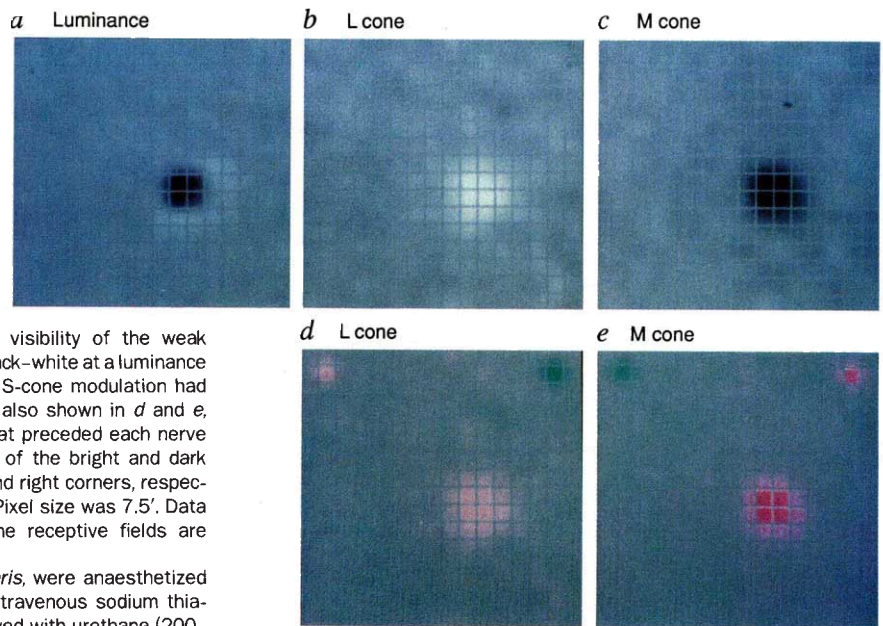
FIG. 1 Receptive field models for red-green selective parvocellular neurons, in this case a red-on neuron. In the cone-specific surround model^{2,3} (a), both centre and surround receive segregated input from a single cone type, here an L on-centre and an M off-surround (L+ and M-). Colour selectivity is determined actively by both of the specifically wired mechanisms. Such a model would predict that the receptive field mapped with a luminance stimulus would have an antagonistic centre-surround organization (b). The receptive field mapped with an L-cone stimulus would show an equally strong centre but no surround (c). The M-cone receptive field would show a response due entirely to the surround mechanism (d). In the mixed surround model⁴⁻⁶ (e), the receptive field centre receives input from one cone type, here represented as an on-response to L cones. The surround, however, receives antagonistic off-input from all cone types at random (L- and M-, also S- in some variants). Colour selectivity in this model is conferred primarily by the sign and cone type of the centre mechanism. The receptive fields measured with luminance (f) or M-cone isolating (h) stimuli would be similar to those predicted for the cone-specific surround model. The critical difference is that the receptive field due to the centre mechanism cone type, here L cones, would show centre-surround antagonism (g), although somewhat weaker than that seen in e. With two exceptions, we never encountered such antagonism in the our cone-isolating receptive fields.

* Present address: Laboratory of Neurobiology, The Rockefeller University, 1230 York Avenue, New York, New York 10021, USA.

FIG. 2 Receptive field of green-off/red-on Type I parvocellular neuron at a stimulus-response interval of 59 ms. The background grid was not present in the stimulus but is used here to indicate the size of the stimulus pixels and the sign of response. The pixel luminances indicate the sign and magnitude of neural excitation. Regions lighter than the grid evoked responses to a lighter stimulus, darker regions to a dark stimulus. The magnitude of the contrast between pixel and grid is proportional to neuronal excitation in impulses per second. Peak responsivities were -7.7 , 10.8 and -16.35 impulses per second per unit contrast. Values displayed in *a* were doubled to enhance visibility of the weak black-white response field. The input stimuli were: black-white at a luminance contrast of 1.0 (*a*); L isolating (*b*); M isolating (*c*). S-cone modulation had no effect on this neuron. L- and M-cone fields are also shown in *d* and *e*, respectively, as the average chromatic stimulus that preceded each nerve impulse. *d* and *e* also show smoothed examples of the bright and dark phases of the cone isolating stimuli, in upper left and right corners, respectively. This neuron was located at 11° eccentricity. Pixel size was $7.5'$. Data were collected for 16 min per receptive field. The receptive fields are smoothed by a function that falls to 5% in 1 pixel.

METHODS. Cynomolgus monkeys, *Macaca fascicularis*, were anaesthetized with ketamine (10 mg kg^{-1} , intramuscularly) and intravenous sodium thiamylal (25 mg kg^{-1} , supplemented as needed), followed with urethane (200 – 300 mg kg^{-1} , every 12 h) and gallamine triethiodide paralysis (20 – 40 mg h^{-1})¹⁸. Action potentials of single neurons in the LGN were recorded with glass-coated tungsten microelectrodes. The visual stimuli were created by an instrument designed and built in our laboratory²³, and were displayed on a Tektronix 690SR colour monitor. The stimulus consisted of a 16×16 grid of square pixels. For every frame of the stimulus (every 14.8 ms), each pixel was assigned one of two values, according to a binary temporal signal, known as a maximal-length shift register sequence, or *m*-sequence⁹. Successive pixel locations used the same *m*-sequence with a relative shift of 256 frames (4 s). A complete *m*-sequence had $2^{16} - 1 (= 65,535)$ frames and lasted 16 min. The neuronal spike trains were correlated with the input *m*-sequences by means of the computationally efficient fast *m*-transform

method²⁴. Pixel sizes ranged from 7.5 to 26 mins of arc. Mean luminance was 225 cd m^{-2} in underscan mode. Each pixel varied between two stimulus levels. The L-cone isolating stimuli were red and green that were indistinguishable (metameric) to the M and S cones. The L-cone contrast was $(L_r - L_g)/(L_r + L_g)$, where L_r was the cone excitation of the red phase of the stimulus and L_g the excitation produced by the green. Cone excitations were calculated by measuring spectral energy distributions of the stimuli at 2 nm resolution (with a Photo Research Model 703 Spectroradiometer), multiplying by the Smith-Pokorny cone absorption spectra²⁵, and summing the products. We used similar colour pairs to those that stimulated the M and S cones in isolation. The cone contrasts used were: L, 0.21 ; M, 0.25 ; S, 0.82 .



method²⁴. Pixel sizes ranged from 7.5 to 26 mins of arc. Mean luminance was 225 cd m^{-2} in underscan mode. Each pixel varied between two stimulus levels. The L-cone isolating stimuli were red and green that were indistinguishable (metameric) to the M and S cones. The L-cone contrast was $(L_r - L_g)/(L_r + L_g)$, where L_r was the cone excitation of the red phase of the stimulus and L_g the excitation produced by the green. Cone excitations were calculated by measuring spectral energy distributions of the stimuli at 2 nm resolution (with a Photo Research Model 703 Spectroradiometer), multiplying by the Smith-Pokorny cone absorption spectra²⁵, and summing the products. We used similar colour pairs to those that stimulated the M and S cones in isolation. The cone contrasts used were: L, 0.21 ; M, 0.25 ; S, 0.82 .

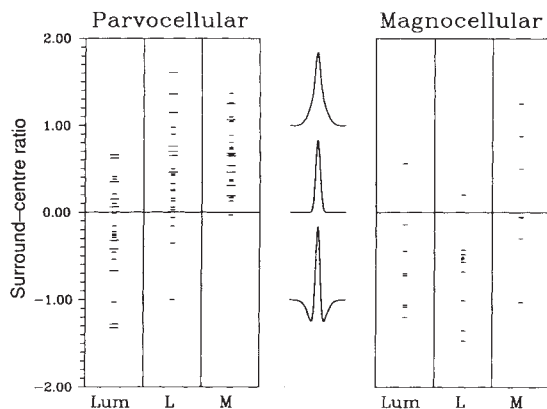


FIG. 3 Scatter plots of surround-centre ratios for 8 magnocellular and 28 red-green opponent parvocellular neurons. The visual fields of parvocellular neurons ranged from 3° to 13° in eccentricity and those of the magnocellular neurons from 3° to 23° . Data were obtained for luminance modulated, L-cone and M-cone isolating stimuli. A negative ratio corresponds to an antagonistic centre-surround organization. In the middle, we show receptive field profiles with an antagonistic centre-surround configuration (ratio = -1.0), a centre only (ratio = 0.0), and an additive centre-surround configuration (ratio = 1.0). To separate the strong centre response from weaker surround effects we arbitrarily assumed that any pixels giving a response of magnitude less than 2 standard deviations above the background, or of opposite sign from the strongest responses, came from the surround. The centre and surround response strengths were taken as the sum of the responses in these respective regions. Surround-centre ratios were calculated separately for the luminance-modulated, L-cone isolating, and M-cone isolating stimuli. Parvocellular neurons with L-cone centres are labelled with short dashes (-), M-cone centre neurons with long dashes (—).

mixed-surround model. Instead we found that centre and surround both receive cone-specific inputs.

For seven of eight magnocellular neurons, both the luminance and L-cone receptive field showed centre-surround antagonism. M-cone receptive fields were frequently non-antagonistic. The stronger L-cone input to magnocellular surrounds, present in both on- and off-centre units, is perhaps a generalization of the type IV on-centre magnocellular neurons with 'tonic red surrounds'¹².

Our results demonstrate that parvocellular neurons, and retinal P cells, receive specific antagonistic inputs from L and M cones. Recent anatomical results suggest that macaque retinal horizontal cells receive non-specific inputs from all cone types^{7,8}. These two sets of results imply that horizontal cells are not responsible for colour opponency. Instead, linear excitatory and

inhibitory inputs to P ganglion cells are mediated by interneurons that receive inputs from a single cone class^{7,20,21}. Because the midget bipolars (one cone per bipolar) extend beyond 10° , these cells could deliver cone-specific signals to amacrine and ganglion cells in the inner retina²². □

Received 19 September 1991; accepted 20 February 1992.

- DeValois, R. L., Morgan, J. C., Polson, M. C., Mead, W. R. & Hull, E. M. *Vision Res.* **14**, 53–67 (1974).
- Wiesel, T. N. & Hubel, D. H. *J. Neurophysiol.* **29**, 1115–1156 (1966).
- Gouras, P. *J. Physiol.* **199**, 533–547 (1968).
- Paulus, W. & Kroger-Paulus, A. *Vision Res.* **23**, 529–540 (1983).
- Shapley, R. & Perry, V. H. *Trends Neurosci.* **9**, 229–235 (1986).
- Lennie, P., Haake, P. W. & Williams, D. R. in *Computational Models of Visual Processing* (eds Landy, M. S. & Movshon, J. A.) 71–82 (MIT, Cambridge, 1991).
- Boynton, B. B., Hopkins, J. M. & Sperling, H. G. *Proc. R. Soc. B* **229**, 345–379 (1987).
- Wassle, H., Boynton, B. B. & Rohrenbeck, J. *Eur. J. Neurosci.* **1**, 421–435 (1991).
- Sutter, E. *Adv. Meth. Physiol. Syst. Modelling* **1**, 303–315 (Univ. Southern California, 1987).

10. Citron, M. C., Kroeker, J. P. & McCann, G. D. *J. Neurophysiol.* **46**, 1161–1176 (1981).
11. Jones, J. P. & Palmer, L. A. *J. Neurophysiol.* **58**, 1187–1211 (1987).
12. Gaska, J. P., Jacobson, L. D., Chen, H.-W. & Pollen, D. A. *Soc. Neurosci. Abstr.* **15**, 1056 (1989).
13. Reid, R. C., Shapley, R. M. & Victor, J. D. *Soc. Neurosci. Abstr.* **15**, 323 (1989).
14. Reid, R. C. & Shapley, R. M. *Invest. Ophthalmol. Vis. Sci.* **31** (suppl.), 429 (1990).
15. Estevez, O. & Spekreijse, H. *Vision Res.* **22**, 681–691 (1982).
16. Gielen, C. C. A. M., van Gisbergen, J. A. M. & Vendrik, A. J. H. *Biol. Cybern.* **44**, 211–221 (1982).
17. Lee, B. B., Martin, P. R. & Valberg, A. *J. Physiol.* **414**, 223–243 (1989).
18. Kaplan, E. & Shapley, R. *Expl. Brain Res.* **55**, 111–116 (1984).
19. Kaplan, E. & Shapley, R. M. *Proc. natn. Acad. Sci. U.S.A.* **83**, 2755–2757 (1986).
20. DeMonasterio, F. M. & Gouras, P. J. *J. Physiol.* **251**, 167–195 (1975).
21. Hubel, D. & Livingstone, M. *Cold Spring Harb. Symp. quant. biol.* **55**, 643–649 (1991).
22. Boycott, B. B. & Wässle, H. *Eur. J. Neurosci.* **5**, 1069–1088 (1991).
23. Milkman, N. et al. *Behav. Res. Meth. Instrum.* **12**, 283–292 (1980).
24. Sutter, E. E. *SIAM J. Comput.* **20**, 686–694 (1992).
25. Smith, V. C. & Pokorny, J. *Vision Res.* **12**, 2059–2071 (1972).

ACKNOWLEDGEMENTS. We thank J. Victor for collaboration in developing the *m*-sequence stimulus, N. Milkman for technical support, Z. Zhang for data processing, and M. Hawken and D. Groszof for help with experiments. This work was supported by the NIH and the Sloan Foundation.

Signal for T-cell differentiation to a CD4 cell lineage is delivered by CD4 transmembrane region and/or cytoplasmic tail

Rho H. Seong, John W. Chamberlain[†] & Jane R. Parnes[‡]

Division of Immunology and Rheumatology, Department of Medicine, Stanford University School of Medicine, Stanford, California 94305-5111, USA

[†] Research Institute, The Hospital for Sick Children, and Department of Immunology, University of Toronto, Toronto, Ontario M5G 1X8, Canada

MATURE T cells express either CD4 or CD8 on their surface. Most helper T cells express CD4, which binds to class II major histocompatibility complex (MHC) proteins, and most cytotoxic T cells express CD8, which binds to class I MHC proteins. In the thymus, mature CD4⁺CD8[−] and CD4[−]CD8⁺ T cells expressing $\alpha\beta$ T-cell antigen receptors (TCR) develop from immature thymocytes through CD4⁺CD8⁺ $\alpha\beta$ TCR⁺ intermediates¹. Experiments using mice transgenic for $\alpha\beta$ TCR^{2–5} suggest that the specificity of the TCR determines the CD4/CD8 phenotype of mature T cells. These results, however, do not indicate how a T cell differentiates into the CD4 or CD8 lineage. Here we show that the CD4 transmembrane region and/or cytoplasmic tail mediates the delivery of a specific signal that directs differentiation of T cells to a CD4 lineage. We generated transgenic mice expressing a hybrid molecule composed of the CD8 α extracellular domains linked to the CD4 transmembrane region and cytoplasmic tail. We predicted that this hybrid molecule would bind to class I MHC proteins through the extracellular domains but deliver the intracellular signals characteristic of CD4. By crossing our transgenic mice with mice expressing a transgenic $\alpha\beta$ TCR specific for a particular antigen plus class I MHC protein, we were able to express the hybrid molecule in developing thymocytes expressing the class I MHC-restricted TCR. Our results show that the signal transduced by the hybrid molecule results in the differentiation of immature thymocytes expressing a class I-restricted TCR into mature T cells expressing CD4.

The CD8/CD4 hybrid DNA construct used for generating transgenic mice is shown in Fig. 1. The resulting β -AprCD8/4⁺ transgenic mice (H-2^d) were crossed with mice (H-2^b) expressing a transgenic $\alpha\beta$ TCR specific for the male (H-Y) antigen in the context of class I H-2D^b MHC antigens⁶ ($\alpha\beta$ TCR⁺ mice) in order to express the hybrid protein in cells expressing a class I-restricted TCR. To examine the effects of the hybrid CD8/CD4 in an H-2^b background (Fig. 2), bone marrow cells from the

$\alpha\beta$ TCR⁺ mice, $\alpha\beta$ TCR⁺ plus β -AprCD8/4⁺ double-transgenic mice, and $\alpha\beta$ TCR⁺ plus β -AL2.1⁺ (expressing a native transgenic CD8 α) double-transgenic mice were transferred to X-irradiated C57BL/6J mice. As shown in Fig. 2, recipients of bone marrow cells from the $\alpha\beta$ TCR⁺ transgenic mice and the $\alpha\beta$ TCR⁺ plus β -AL2.1⁺ double-transgenic mice have a minor proportion of transgenic TCR⁺ cells (T3.70^{hi}) expressing CD4 in their lymph nodes (5–8% of T3.70^{hi} cells are CD4⁺; Fig. 2a and b). In contrast, in recipients of bone marrow cells from the $\alpha\beta$ TCR⁺ plus β -AprCD8/4⁺ double-transgenic mice, about 50% of lymph node cells expressing the transgenic TCR also express CD4 (Fig. 2c). In the total lymph node CD4⁺ cell population of bone marrow recipients from the $\alpha\beta$ TCR⁺ plus β -AprCD8/4⁺ double-transgenic mice, 23% of cells express CD8 (the transgenic CD8/CD4 hybrid $\pm$ endogenous CD8) (Fig. 3a). Of these CD4⁺CD8⁺ cells, only 13% express endogenous CD8 α , as detected by staining for Lyt-2.1 (data not shown), indicating that 87–100% express the hybrid CD8/CD4 transgene. The transgenic $\alpha\beta$ TCR⁺ (T3.70^{hi}) cells expressing CD4 in the $\alpha\beta$ TCR⁺ plus β -AprCD8/4⁺ recipients were found almost exclusively in this population staining for CD8 α (Fig. 3b and c). Again, CD8 α staining detected in this population is due to expression of the hybrid CD8/CD4 molecule on most, if not all, of these cells (see below).

The increased proportion of T3.70^{hi}CD4⁺ cells is also observed in the thymus, implying a commitment of these cells to CD4 expression during thymocyte differentiation (Fig. 4a–c). The number of T3.70^{hi}CD4⁺ cells lacking expression of endogenous CD8 (CD8^{endo−}) (as determined by CD8 β staining) is increased from 1–2% of the T3.70^{hi} cells in the transgenic $\alpha\beta$ TCR⁺ or $\alpha\beta$ TCR⁺ plus β -AL2.1⁺ recipients to at least 4.9% of T3.70^{hi} cells in the $\alpha\beta$ TCR⁺ plus β -AprCD8/4⁺ recipients (Fig. 4a–c). The ratio of CD4⁺CD8^{endo−} to CD4[−]CD8^{endo+} cells is increased 12.5–20 fold in the thymus of the $\alpha\beta$ TCR⁺ plus β -AprCD8/4⁺ recipients (1:2) compared with $\alpha\beta$ TCR⁺ (1:25) or $\alpha\beta$ TCR⁺ plus β -AL2.1⁺ (1:40) recipients. The increase in this ratio in the $\alpha\beta$ TCR⁺ plus β -AprCD8/4⁺ recipients is the result of both the increased proportion of T3.70^{hi}CD4⁺CD8^{endo−} cells and the decreased proportion of T3.70^{hi}CD4[−]CD8^{endo+} cells. These changes are presumably the result of expression of CD4 by some cells which, in the absence of the hybrid CD8/CD4 transgene,

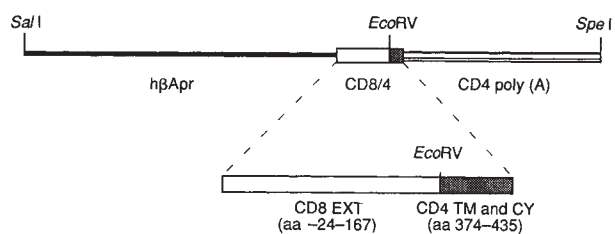
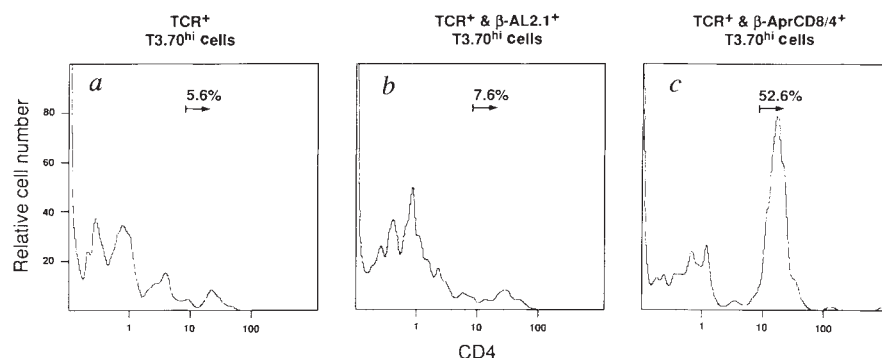


FIG. 1 Structure of the β -AprCD8/4 transgene. The CD8/CD4 complementary DNA hybrid construct has been described¹². The encoded hybrid protein is composed of the signal peptide (amino acids (aa) −24 to −1) and extracellular domains (EXT) (amino acids 1–167) of CD8 α (Lyt-2.2) and the transmembrane region (TM) and cytoplasmic tail (CY) of CD4 (amino acids 374–435). The junctional amino-acid sequence for the hybrid protein is -FACD/ILA- (one-letter code). The hybrid cDNA was placed under the control of human β -actin regulatory sequences¹³. Part of the 3' untranslated region and the 3' flanking region (2.7-kb *Bgl*III–*Eco*RI fragment) of the mouse CD4 gene¹⁴ were inserted at the 3' end of the hybrid cDNA to provide a poly(A) signal. An 8-kb *Sal*I–*Spe*I fragment of the CD8/CD4 hybrid gene construct was isolated and used for production of transgenic mice by microinjection into C3H/J $\times$ DBA/2J F2 embryos (Lyt-2.1) as described (R.H.S. et al., manuscript submitted). Two transgenic lines were established by backcrossing founder mice to DBA/2J partners. The mice used in this study were progeny after two backcrosses. The resultant transgenic mice expressed the CD8/CD4 protein on all lymphoid cells as determined by cytofluorometry after staining with an anti-CD8 α monoclonal antibody specific for Lyt-2.2 (ref. 15).

[‡] To whom correspondence should be addressed.

FIG. 2 Surface expression of CD4 by transgenic $\alpha\beta$ TCR⁺ lymph node cells of C57BL/6J female recipients of bone marrow cells from the $\alpha\beta$ TCR⁺ (a), $\alpha\beta$ TCR⁺ plus β -AL2.1⁺ (b), and $\alpha\beta$ TCR⁺ plus β -AprCD8/4⁺ (c) transgenic mice. The β -AprCD8/4⁺ transgenic mice (DBA/2J background; H-2^d) were crossed with the $\alpha\beta$ TCR⁺ transgenic mice⁶ (C57BL/6J background; H-2^b). Bone marrow cells from female offspring expressing the transgenic $\alpha\beta$ TCR alone ($\alpha\beta$ TCR⁺; DBA/2J $\times$ C57BL/6J) or both the transgenic $\alpha\beta$ TCR and the hybrid CD8/CD4 ($\alpha\beta$ TCR⁺ plus β -AprCD8/4⁺; DBA/2J $\times$ C57BL/6J) were transferred into X-irradiated female C57BL/6J mice. In addition, bone marrow cells from female mice expressing both the transgenic $\alpha\beta$ TCR and a native CD8 α transgene (driven by the human β -actin regulatory region) from previously described β -AL2.1 transgenic mice (C57BL/6J background) (R.H.S. *et al.*, manuscript submitted) were also transferred into X-irradiated female C57BL/6J mice. Transgenic TCR expression was monitored with the monoclonal antibody T3.70, which is specific for the transgenic α -chain but binds 10-fold better when it is associated with the transgenic β -chain (T3.70^{hi}) as opposed to other β -chains¹⁶. In $\alpha\beta$ TCR female transgenic mice of the H-2^b haplotype, most lymph node T cells expressing high levels of both the transgenic TCR α - and β -chain are CD4⁺CD8⁺ as a result of positive selection¹⁷. The CD4⁺CD8⁺ lymph node T cells in these mice result from incomplete allelic exclusion of the TCR α -chain^{16,17}. But in $\alpha\beta$ TCR female transgenic mice of the non-selecting H-2^d haplotype, a few of the CD4⁺CD8⁺ lymph node T cells (<10% of lymph node T cells) express high levels of the transgenic TCR α -chain¹⁸. To avoid any selection for CD4⁺CD8⁺ T cells expressing the transgenic TCR α - and β -chains as a result of the H-2^d background¹⁸ contributed by the β -AprCD8/4⁺ transgenic mice, and as thymic epithelial cells rather than bone marrow-derived cells are responsible for positive selection^{17,19-22}, bone marrow transplants were done to examine the effects



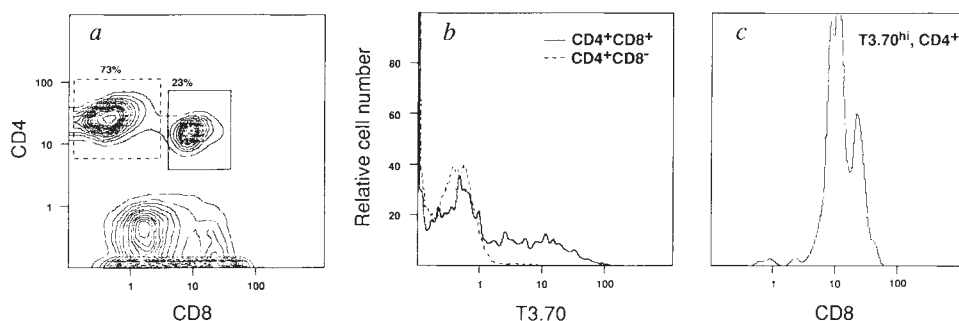
of the hybrid CD8/CD4 in an H-2^b background. The plotted histograms correspond to the levels of CD4 expression on cells gated for high expression of transgenic $\alpha\beta$ TCR (T3.70^{hi}). The percentages of T3.70^{hi} cells expressing CD4 are shown. These plots show one of three similar results from independent experiments.

METHODS. The transgenic β -AprCD8/4⁺ or β -AL2.1⁺ mice were crossed with mice expressing the transgenic $\alpha\beta$ TCR specific for H-Y+H-2D^b (ref. 18). Bone marrow cells (4×10^6 cells) from female progeny expressing the $\alpha\beta$ TCR, the $\alpha\beta$ TCR plus β -AL2.1, or the $\alpha\beta$ TCR plus β -AprCD8/4 transgenes were transferred to lethally irradiated (900 R) C57BL/6J female mice. Three to five weeks after the transfer, single-cell suspensions of pooled lymph nodes from the bone marrow recipients were prepared and stained with biotin-conjugated T3.70, followed by allophycocyanin-conjugated streptavidin and phycoerythrin-conjugated anti-CD4 (both Becton-Dickinson). A total of 20,000 viable cells per sample were analysed in each sample on a modified FACS II using Facscan Research Software Program (Becton-Dickinson).

would have been T3.70^{hi}CD4⁺CD8^{endo+}. This interpretation is suggested by the higher proportion of CD4⁺CD8⁺T3.70^{hi} cells in the thymus of the $\alpha\beta$ TCR⁺ plus β -AprCD8/4⁺ recipients (about 56%) than in the other recipients (about 26–28%) (Fig. 4a–c). This is further supported by the observation that about 30% of T3.70^{hi}CD4⁺ lymph node cells also express endogenous CD8 β in the $\alpha\beta$ TCR⁺ plus β -AprCD8/4⁺ recipients (Fig. 4f). The marked reduction in the percentage of T3.70^{hi}CD4⁺CD8^{endo+} cells in the $\alpha\beta$ TCR⁺ plus β -AprCD8/4⁺ recipients (8.6%; Fig. 4c) compared with that in the other recipients (46.6% and 37.8%; Fig. 4a and b) suggests that T3.70^{hi}CD4⁺(CD8^{endo+} or CD8^{endo-}) cells in the $\alpha\beta$ TCR⁺ plus β -AprCD8/4⁺ recipients originate from cells that would have otherwise been in the

T3.70^{hi}CD4⁺CD8^{endo+} subpopulation. Thus the T3.70^{hi}CD4⁺ cells in the $\alpha\beta$ TCR⁺ plus β -AprCD8/4⁺ recipients seem to be instructed to maintain expression of CD4 instead of becoming T3.70^{hi}CD4⁺CD8^{endo+}. It is possible that the development of T3.70^{hi}CD4⁺CD8^{endo+} cells is severely hindered by expression of the hybrid protein on developing thymocytes because of the higher affinity of its CD4 cytoplasmic tail than that of endogenous CD8 for the tyrosine kinase p56^{lck} (refs 7–9). However, the greatly increased proportion of T3.70^{hi}CD4⁺CD8⁺ cells in the thymus of the $\alpha\beta$ TCR⁺ plus β -AprCD8/4⁺ recipients compared with other recipients (Fig. 4a–c) argues against this possibility. Notably, the expression of the transgenic native CD8 α in the $\alpha\beta$ TCR⁺ plus β -AL2.1⁺

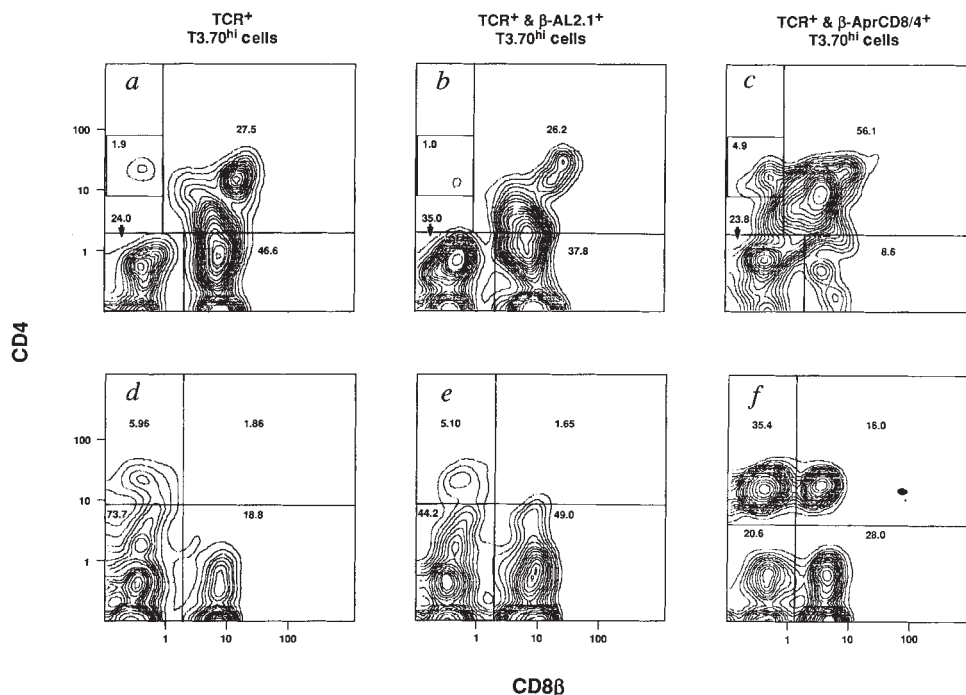
FIG. 3 CD4⁺T3.70^{hi} cells in C57BL/6J female recipients of $\alpha\beta$ TCR⁺ plus β -AprCD8/4⁺ bone marrow are found primarily in the CD4⁺CD8 α ⁺ population and not in CD4⁺CD8 α ⁻ population. Expression of CD4 and CD8 α on total lymph node cells from the $\alpha\beta$ TCR⁺ plus β -AprCD8/4⁺ bone marrow recipient C57BL/6J mice is presented as a density plot (a). Cells were then gated for the expression of CD8 α and presented as a histogram for the expression of the transgenic TCR (b). The level of the transgenic $\alpha\beta$ TCR on CD4⁺ cells expressing the transgenic CD8/CD4 hybrid ($\pm$ endogenous CD8) is shown as a solid line and that of the CD4⁺ cells not expressing the CD8/CD4 hybrid (or endogenous CD8) is shown as a dotted line (b). Alternatively, the cells gated for the expression of both CD4 and the transgenic $\alpha\beta$ TCR are presented as a histogram for the expression of CD8 α (c). Note that the CD8 portion of the CD8/CD4 hybrid molecule is Lyt-2.2 whereas endogenous CD8 consists of both Lyt-2.1 and Lyt-2.2. We can detect cells expressing endogenous CD8 by the presence of Lyt-2.1 and the presence of the CD8 β chain, but if a cell does express endogenous CD8, we cannot be certain that it also expresses the hybrid molecule on the surface by FACS analysis. Even though all lymphoid cells express the transgenic CD8 α on the surface in both β -AL2.1⁺ transgenic mice and β -AprCD8/4⁺ transgenic mice, not all cells express the transgenic CD8 α after bone marrow



transfer. Most thymocytes in the bone marrow recipients express the transgenic TCR, suggesting that most thymocytes are from the transferred bone marrow cells. The lymph node cells that do not express the transgenes may be residual host T cells that survived the X-irradiation and/or could be cells from the transferred bone marrow which for some reason do not express the transgenes.

METHODS. For details, see Fig. 2. Single cell suspensions of pooled lymph nodes from C57BL/6J female recipients of bone marrow from $\alpha\beta$ TCR⁺ plus β -AprCD8/4⁺ mice were stained with biotin-conjugated T3.70 followed by allophycocyanin-conjugated streptavidin, fluorescein isothiocyanate-conjugated anti-CD8 α and phycoerythrin-conjugated anti-CD4 (all from Becton-Dickinson).

FIG. 4 Staining of thymocytes and lymph node cells from female C57BL/6J recipients of bone marrow from the $\alpha\beta$ TCR⁺ (a and d), $\alpha\beta$ TCR⁺ plus β -AL2.1⁺ (b and e), and $\alpha\beta$ TCR⁺ plus β -AprCD8/4⁺ (c and f) transgenic mice. Surface expression of CD4 and endogenous CD8 β on thymocytes from the $\alpha\beta$ TCR⁺ (a), $\alpha\beta$ TCR⁺ plus β -AL2.1⁺ (b), or $\alpha\beta$ TCR⁺ plus β -AprCD8/4⁺ (c) recipients is presented as density plots with the cells gated for the 5% highest expression of the transgenic $\alpha\beta$ TCR detected with monoclonal antibody T3.70. The skewing of the CD4⁺:CD8^{endo+} ratio in the $\alpha\beta$ TCR⁺ transgenic mice (1.9% versus 46.6%; ratio of about 1:25) (a) compared with that of a normal mouse (2:1) is a result of positive selection of this TCR on CD8^{endo+} cells¹⁷. The CD4⁺:CD8^{endo+} ratio in $\alpha\beta$ TCR⁺ plus β -AL2.1⁺ transgenic mice (1.0% versus 37.8%) is even further skewed (b) compared with the $\alpha\beta$ TCR⁺ transgenic mice, as previously described¹⁰ in CD8 plus $\alpha\beta$ TCR double-transgenic mice. Surface expression of CD4 and endogenous CD8 β is also presented for lymph node cells expressing the transgenic TCR (T3.70^{hi}) from the $\alpha\beta$ TCR⁺ (d), $\alpha\beta$ TCR⁺ plus β -AL2.1⁺ (e), and $\alpha\beta$ TCR⁺ plus β -AprCD8/4⁺ (f) recipients. The percentages of cells corresponding to each sector are indicated. It should be noted that continued expression of endogenous CD8 β on about 30% of the CD4⁺T3.70^{hi} lymph node cells in the $\alpha\beta$ TCR⁺ plus β -AprCD8/4⁺ recipients (f) implies that the hybrid CD8/CD4 transgene is expressed on at least the remaining 70%, if not all, of these cells (all were shown to be CD8⁺; Fig. 2b and c). Three independent experiments gave



similar results.

METHODS. For details, see Fig. 2. Thymocytes and lymph node cells from the bone marrow cell recipients were first stained with monoclonal antibody 53-5, which is specific for mouse CD8 β (ref. 23) followed by Texas red-conjugated mouse anti-rat antibody. After thorough washing, the cells were stained with biotin-conjugated T3.70 followed by allophycocyanin-conjugated avidin and phycoerythrin-conjugated anti-CD4 (Becton-Dickinson).

recipients did not increase T3.70^{hi}CD4⁺ cells (Figs 2b, 4b and e) compared with the $\alpha\beta$ TCR⁺ recipients (Figs 2a and 4a, d). Similar results have been reported^{10,11} and are considered to be inconsistent with a strict stochastic/selective model for T-cell development. Our results further indicate that the transmembrane region and/or cytoplasmic tail of CD4 delivers a specific signal for thymocyte differentiation which leads to maintenance of CD4 expression.

It is surprising that some of the T3.70^{hi} cells in lymph node express both CD4 and CD8 β (Fig. 4f). Most CD4⁺CD8 β ⁺ cells, which are found only in the $\alpha\beta$ TCR⁺ plus β -AprCD8/4⁺ recipients, also expressed endogenous CD8 α (Lyt-2.1) (data not shown). These T3.70^{hi}CD4⁺CD8^{endo+} cells do not seem to originate from cells that were pre-committed to the CD8 lineage and then rescued to express CD4 by a signal through the CD4 transmembrane and cytoplasmic domain of the CD8/CD4 hybrid, because cells pre-committed to the CD4 lineage and rescued by the transgene should then have also been detected in the $\alpha\beta$ TCR⁺ plus β -AL2.1⁺ recipients (Fig. 4e). It is most

likely that the T3.70^{hi}CD4⁺CD8^{endo+} peripheral cells result from competition between the transgenic hybrid CD8/CD4 and endogenous CD8 for binding to the same class I MHC as the TCR at the time of selection. Some cells may have received independent signals from the hybrid and endogenous CD8 molecules to maintain both CD4 and CD8^{endo} expression, and these signals might be dominant over a separate signal to turn off one of the molecules. Alternatively, expression of CD4 and CD8 may be temporary in immature thymocytes, and continued expression of these proteins could be established by a signal delivered by the corresponding molecule during positive selection.

Taken together, our results indicate that a specific signal for differentiation of an immature thymocyte to the CD4 or CD8 lineage is delivered by each molecule through its transmembrane region and/or cytoplasmic tail, respectively. These results demonstrate that the CD4 and CD8 coreceptor molecules play an active part in signalling and directing the developmental pattern of thymocytes. □

Received 24 October 1991; accepted 28 February 1992.

- Fowlkes, B. J. & Pardoll, D. M. *Adv. Immun.* **44**, 207-264 (1989).
- Teh, H.-S. *et al. Nature* **335**, 229-233 (1988).
- Sha, W. C. *et al. Nature* **335**, 271-274 (1988).
- Berg, L. J. *et al. Cell* **58**, 1035-1046 (1989).
- Kaye, J. *et al. Nature* **341**, 746-749 (1989).
- Kisielow, P., Blüthman, H., Staerz, U. D., Steinmetz, M. & von Boehmer, H. *Nature* **333**, 742-746 (1988).
- Veillette, A., Bookman, M. A., Horak, E. M. & Bolen, J. B. *Cell* **55**, 301-308 (1988).
- Veillette, A., Zúñiga-Pflücker, J. C., Bolen, J. B. & Kruisbeek, A. M. *J. exp. Med.* **170**, 1671-1680 (1989).
- Luo, K. & Sefton, B. M. *Molec. cell. Biol.* **10**, 5305-5313 (1990).
- Robey, E. A. *et al. Cell* **64**, 99-107 (1991).
- Borgulya, P., Kishi, H., Müller, U., Kirberg, J. & von Boehmer, H. *EMBO J.* **10**, 913-918 (1991).
- Zamoyiska, R. *et al. Nature* **342**, 278-281 (1981).
- Gunning, P., Leavitt, J., Muscat, G. Ng, S. & Kedes, L. *Proc. natn. Acad. Sci. U.S.A.* **84**, 4831-4835 (1987).

- Gorman, S. D., Tourville, B. & Parnes, J. R. *Proc. natn. Acad. Sci. U.S.A.* **84**, 7644-7648 (1987).
- Sarmiento, M. *et al. Immunol. Rev.* **68**, 135-169 (1982).
- Teh, H.-S., Kishi, H., Scott, B. & von Boehmer, H. *J. exp. Med.* **169**, 795-806 (1989).
- Kisielow, P., Teh, H.-S., Blüthman, H. & von Boehmer, H. *Nature* **335**, 730-733 (1988).
- Teh, H.-S. *et al. Nature* **335**, 229-233 (1988).
- Bevan, M. J. *Nature* **269**, 417-418 (1977).
- Zinkernagel, R. H. *et al. J. exp. Med.* **147**, 882-896 (1978).
- von Boehmer, H., Haas, W. & Jerne, N. K. *Proc. natn. Acad. Sci. U.S.A.* **75**, 2434-2442 (1978).
- Kappler, J. & Marrack, P. *J. exp. Med.* **148**, 1510-1522 (1978).
- Ledbetter, J. A., Seaman, W., Tsu, T. & Herzenberg, L. A. *J. exp. Med.* **153**, 1503-1516 (1981).

ACKNOWLEDGEMENTS. We thank H. von Boehmer and H.-S. Teh for the $\alpha\beta$ TCR⁺ transgenic mice and monoclonal antibody T3.70, and P. Sullivan, S. Phillips, J. Chen and G. D. Frank for technical assistance. This work was supported by the NIH and a gift from the Baxter Foundation. J.R.P. is an Established Investigator of the American Heart Association. J.W.C. is a Medical Research Council of Canada Scholar.

Nonsense mutation in the glucokinase gene causes early-onset non-insulin-dependent diabetes mellitus

N. Vionnet^{*†‡}, M. Stoffel^{*‡}, J. Takeda^{*‡}, K. Yasuda^{*},
G. I. Bell^{*§}, H. Zouali^{†||}, S. Lesage^{†||}, G. Velho^{†||},
F. Iris[†], Ph. Passa[¶], Ph. Froguel^{†§¶} & D. Cohen[†]

^{*} Howard Hughes Medical Institute, Departments of Biochemistry & Molecular Biology and Medicine, The University of Chicago, 5841 South Maryland Avenue, MC1028, Chicago, Illinois 60637, USA
[†] Centre d'Etude du Polymorphisme Humain, 27 rue Juliette Dodu, and
^{||} G  n  thon, 1 rue de l'Internationale, 91000 Evry, France
[¶] Service d'Endocrinologie, H  pital Saint-Louis, 75010 Paris, France
[‡] These authors contributed equally to this letter
[§] To whom correspondence should be addressed

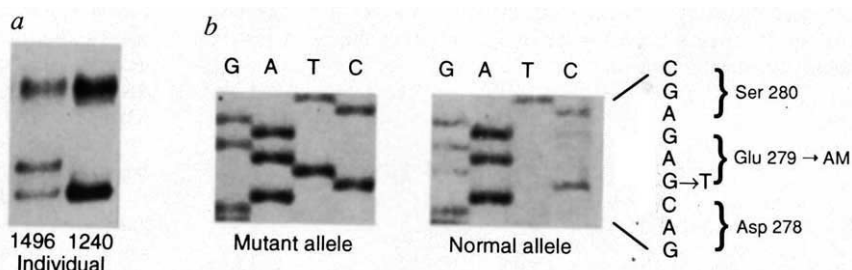
MATURITY-ONSET diabetes of the young (MODY) is a form of non-insulin-dependent (type 2) diabetes mellitus (NIDDM) which is characterized by an early age at onset and an autosomal dominant mode of inheritance¹. Except for these features, the clinical characteristics of patients with MODY are similar to those with the more common late-onset form(s) of NIDDM. Previously² we observed tight linkage between DNA polymorphisms in the glucokinase gene on the short arm of chromosome 7 and NIDDM in a cohort of sixteen French families having MODY. Glucokinase is an enzyme that catalyses the formation of glucose-6-phosphate from glucose and may be involved in the regulation of insulin secretion and integration of hepatic intermediary metabolism³. Because the glucokinase gene was a candidate for the site of the genetic lesion in these families, we scanned this gene for mutations. Here we report the identification of a nonsense mutation in the gene encoding glucokinase and its linkage with early-onset diabetes in one family. To our knowledge, this result is the first evidence implicating a mutation in a gene involved in glucose metabolism in the pathogenesis of NIDDM.

The gene encoding the enzyme glucokinase in humans spans a region of >20 kilobases (kb) and consists of 12 exons (M.S., J.T., N.V. and G.I.B., manuscript in preparation). The exon-intron organization of the human glucokinase gene is similar to that of the rat⁴. The regions of the exons were amplified *in vitro*

using the polymerase chain reaction (PCR)⁵ and were scanned for mutations using the technique of single-strand conformational polymorphism (SSCP)^{6,7}. We investigated two unrelated non-diabetic individuals and an affected member from each of five families (F8, F28, F51, F85 and F393) showing tight linkage between microsatellite DNA polymorphisms in the glucokinase gene and MODY (ref. 2). SSCP analysis of exon 7 in family F8 revealed a pattern in a patient with MODY consistent with the presence of a nucleotide substitution in one allele (Fig. 1a). This pattern was not seen in affected individuals from the other four families. Amplification and direct sequencing of exon 7 in individual 1496 of family F8 revealed a single nucleotide change in this exon. There was a G → T base substitution in codon 279 (using the amino-acid numbering of the β -cell isoform of human glucokinase (this protein is 465 amino acids long)) which changes GAG (glutamic acid) to TAG, an amber termination codon. The presence of this mutation was confirmed by subcloning the PCR product into M13mp18 and determining its sequence (Fig. 1b). SSCP analysis of 34 members of kindred F8, 15 of whom have MODY (Fig. 2), showed that this mutation was only present in individuals with MODY. The presence of the nonsense mutation was also confirmed by direct sequencing of the PCR product in seven other members of kindred F8. In addition, the sequence of exon 7 was normal in three non-diabetic members. As noted previously², there is a patient in this family (individual number 50 in Fig. 2) who has NIDDM but does not show linkage with the DNA polymorphisms in the glucokinase gene that are associated with MODY in this pedigree. Moreover, she does not have the nonsense mutation in codon 279. We consider that this patient has another form of NIDDM as she is obese and had a later age at onset (52 years) of NIDDM than other members of the family with MODY. Thus, there are two forms of NIDDM present in kindred F8. Such heterogeneity has been noted in other families with MODY⁸. The linkage between the nonsense mutation and MODY in family F8 is highly significant. Using the LINKAGE programs⁹, assuming a frequency of the nonsense mutation of 0.0001 in the general population and considering individual 50 as unknown with respect to MODY and all non-diabetic individuals as being normal, a test of no linkage versus linkage, with the assumption of complete linkage disequilibrium between the mutation and the disease allele, gave a test statistic of 10.0 on the lod score scale. SSCP analysis of 34 unrelated non-diabetic subjects and 55 unrelated members of the French families did not show the mutant pattern, implying that the

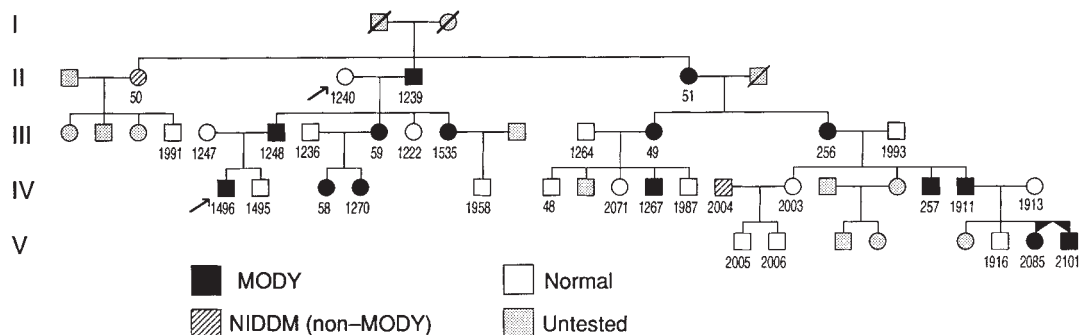
FIG. 1 a, SSCP analysis of exon 7 of the glucokinase gene from a member of kindred F8 with MODY (1496) and a normal relative (1240). b, Sequence of exon 7 in the region of codon 279. The mutant allele is shown on the left and the normal allele on the right. AM, Amber mutation.

METHODS. DNA (100 ng) was digested with *EcoRI* for 1 h and then precipitated with ethanol. This DNA was mixed with 10 pmol of the human glucokinase exon 7-specific primers, hGK7a (upstream, 5'-AGTGCGCTCTCGCTGACAG-3') and hGK7b (downstream, 5'-CATCTGCCGCTG-CACCAGAG-3') using GeneAmp (Perkin Elmer Cetus) reagents and GeneAmp PCR System 9600; the MgCl₂ concentration was 1.5 mM. For SSCP, the reaction volume was 10 μ l and included 0.5 μ Ci of α -³²P-labelled dCTP (Amersham). The PCR conditions were: initial denaturation for 5 min at 94 °C, then 30 cycles of denaturation at 94 °C for 1 min, and annealing and extension at 65 °C for 2 min, with a final extension at 65 °C for 7 min. The amplified PCR product was 285 base pairs (bp) and contained exon 7 (184 bp) and 41 and 60 bp of the flanking introns, including primers. Samples were processed for SSCP analysis according to refs 6 and 7. Fragments were separated on a 40-cm 5% polyacrylamide (24:1 acrylamide:bis) gel run in 1 × Tris-borate buffer at room temperature without glycerol. The gel was run at 6 watts for 12 h, after which the xylene cyanol dye had run the full length of the gel. The gel was exposed to X-ray film overnight. To amplify



exon 7 for sequencing, a reaction volume of 100 μ l was used and the PCR conditions were as above, except 35 cycles were used. After PCR, unincorporated nucleotides and primers were removed using a Centricon-100 microconcentrator (Amicon). The 285-bp PCR product was cloned into the *HincII* site of M13mp18 DNA and sequenced using a Sequenase DNA sequencing kit (USB, Cleveland). PCR products of exon 7 were sequenced directly using a DNA cycle sequencing system (GIBCO-BRL) and primers hGK7a and hGK-22 (5'-CATGTGCGTCAATACCGAGT-3'), or using these primers with an Applied Biosystems DNA sequencer (model 373A); dideoxy-cycle-sequencing reactions (25 cycles) were in tubes containing 80–120 ng of the PCR product using fluorescence-labelled dideoxy terminators as described by Applied Biosystems.

FIG. 2 Partial pedigree of kindred F8; 34 individuals were studied and 17 of these have diabetes mellitus. With the exception of individual number 50, none of the diabetic individuals is obese (that is, body mass index $>26 \text{ kg m}^{-2}$). The mean age of apparent onset of diabetes was 6 and 16 years of age in generations 5 and 4, respectively. Most of the affected individuals in this family have mild fasting hyperglycemia with blood glucose values ranging from 6.7–9.0 mM. Two individuals are at present being treated with sulphonylurea drugs (individuals 59 and 1239; the patient with NIDDM (individual 50) is also being treated with sulphonylurea drugs) and one with insulin (individual 1267). The fasting and 2 h-post-glucose-load insulin levels are not different between affected and non-affected individuals. The form of mild diabetes seen in this family seems to be quite frequent among MODY patients in France¹². The CEPH (Centre d'Etude du Polymorphisme Humain) identification numbers of the



subjects studied are shown. Individuals considered to have MODY using the diagnostic criteria described in ref. 2 are indicated by black symbols. All of these individuals were tested for the nonsense mutation at codon 279 in exon 7 by SSCP or direct-sequencing strategies and, with the exception of individual 50 (this patient is believed to have a form of NIDDM other than MODY, see discussion in text), all were heterozygous for this mutation. The affection status of the members of this family is noted. Arrows indicate the two individuals from this family who were screened using SSCP (Fig. 1a).

nonsense mutation in the glucokinase gene is not a common polymorphism.

MODY has an autosomal dominant mode of inheritance. The molecular mechanism by which mutations in glucokinase may lead to the development of glucose intolerance and diabetes mellitus is unknown. As glucokinase appears to be a monomer and there is no evidence for dimerization¹⁰, it is unlikely that the nonsense peptide binds to the functional protein that is the product of the normal allele and acts as a dominant negative regulator of glucokinase activity. It is more likely that levels of glucokinase activity in the β -cell are critical in determining the threshold at which the β -cell secretes insulin in response to changes in glucose concentration. In this regard, it has been suggested that a modest 15% decrease in glucokinase activity might shift the set point for glucose-induced insulin secretion from 5 to 6 mM (ref. 11).

We are continuing to screen MODY families using SSCP and direct-sequencing strategies for other mutations in the glucokinase gene. In addition, we are screening NIDDM patients having the more common late-onset form of this disorder to determine whether genetic variation in glucokinase also contributes to its development. If mutations in glucokinase can contribute to the development of NIDDM, then acquired changes resulting in decreased levels of glucokinase may have the same effect; glucokinase may therefore be a target for therapeutic intervention. □

Actin-dependent organelle movement in squid axoplasm

Sergei A. Kuznetsov*†‡, George M. Langford*§
& Dieter G. Weiss*†||

* Marine Biological Laboratory, Woods Hole, Massachusetts 02543, USA

† Institut für Zoologie, Technische Universität München, Lichtenbergstrasse 4, D-W8046 Garching, Germany

‡ Department of Molecular Biology, Faculty of Biology, Moscow State University, Moscow 119899, Russia

§ Department of Biological Sciences, Dartmouth College, Hanover, New Hampshire 03755-3576, USA

STUDIES of organelle movement in axoplasm extruded from the squid giant axon have led to the basic discoveries of microtubule-dependent organelle motility^{1–3} and the characterization of the microtubule-based motor proteins kinesin and cytoplasmic dynein^{4,5}. Rapid organelle movement in higher animal cells, especially in neurons, is considered to be microtubule-based. The role of actin filaments, which are also abundant in axonal cytoplasm^{6,7}, has remained unclear. The inhibition of organelle movement in axoplasm by actin-binding proteins^{8–11} such as DNase I, gelsolin and synapsin I has been attributed to their ability to disorganize the microtubule domains where most of the actin-filaments are located⁷. Here we provide evidence of a new type of organelle movement in squid axoplasm which is independent of both microtubules and microtubule-based motors. This movement is ATP-dependent, unidirectional, actin-dependent, and probably generated by a myosin-like motor. These results demonstrate that an actomyosin-like mechanism can be directly involved in the generation of rapid organelle transport in nerve cells.

To study the role of actin filaments in fast axonal transport, we developed a procedure that allowed actin filaments to form an extensive network near the bulk axoplasm extruded from squid giant axon. Axoplasm was extruded into dissociation buffer and the distribution of actin filaments on the glass surface was analysed by perfusing the unfixed preparations with rhodamine-phalloidin¹². We observed a fluorescent filament system inside the bulk axoplasm, but also an extensive filamentous network that formed at the periphery of axoplasm fragments and extended to a distance of more than 100 μm (Fig. 1a). The

Received 17 February; accepted 25 March 1992.

1. Fajans, S. S. *Diabetes/Metabol. Rev.* **5**, 579–606 (1989).
2. Froguel, Ph. *et al. Nature* **356**, 162–164 (1992).
3. Matschinsky, F. M. *Diabetes* **39**, 647–652 (1990).
4. Magnuson, M. A. *Diabetes* **39**, 523–527 (1990).
5. Sakai, R. J. *et al. Science* **239**, 487–491 (1988).
6. Orita, M., Iwahana, H., Kanazawa, H., Hayashi, K. & Sekiya, T. *Proc. natn. Acad. Sci. U.S.A.* **86**, 2766–2770 (1989).
7. Orita, M., Suzuki, Y., Sekiya, T. & Hayashi, K. *Genomics* **5**, 874–879 (1989).
8. Bell, G. I. *et al. Proc. natn. Acad. Sci. U.S.A.* **88**, 1484–1488 (1991).
9. Lathrop, G. M., Lalouel, J. M., Julier, C. & Ott, J. *Am. J. hum. Genet.* **37**, 482–498 (1985).
10. Holroyde, M. J. *et al. Biochem. J.* **153**, 363–373 (1976).
11. Meglasson, M. D. & Matschinsky, F. M. *Am. J. physiol.* **246**, E1–E13 (1984).
12. Lestrade, H. *et al. Arch. Fr. Pediatr.* **46**, 19–23 (1989).

ACKNOWLEDGEMENTS. We thank the patients for their cooperation; H. Lestrade, CNP-Assurances, the ATP, and Boehringer Mannheim for help in collecting the families; and M. Lathrop for his help in calculating the test statistic. M.S. and K.Y. were supported by fellowships from the Deutsche Forschungsgemeinschaft and a Mentor-based Fellowship from the American Diabetes Association, respectively. These studies were supported by the Howard Hughes Medical Institute, NIH, Association Française contre les Myopathies (through the Genethon program), the Assistance Publique-Hôpitaux de Paris and the French Ministry for Research and Technology.

|| To whom correspondence should be addressed in Germany.

TABLE 1 Conditions for actin filament network formation and for organelle movement on actin networks and free microtubules

Treatment	Concentration	ATP (mM)	Actin filaments		Microtubules	
			Motility*	Network formation†	Motility*	Presence of free microtubules‡
Control	—	2	+	+	+	+
Control	—	—	+	+	+	+
Cytochalasin B	10 μ M	2	—	—	+	+
Nocodazole	100 μ M	2	+	+	—	—
Apyrase	5 U ml ⁻¹	—	—	±	—	+
Hexokinase + glucose	10 U ml ⁻¹ + 5 mM	—	—	+	—	+
AMP-PNP	4 mM	2	+	+	±	+
	4 mM	—	+	+	—	+
Vanadate	10 μ M	2	+	+	+	+
	100 μ M	2	+	+	±	+
EDTA	5 mM	2	—	—	—	±
EGTA	5 mM	2	+	+	+	+

Axoplasm was extruded into buffer containing 0.4 μ M rhodamine-phalloidin and observed for one hour by both fluorescence microscopy (for network formation) and AVEC-DIC microscopy (for organelle motility). Inhibitors were added to the buffer containing rhodamine-phalloidin with or without ATP. For the EDTA experiments, Mg²⁺ and Ca²⁺ were omitted from the buffer, and for the EGTA experiments Ca²⁺ was omitted. With all treatments, half of the axoplasm from the same axon was used as control for the presence of motility.

* The motile activity is expressed as the number (the average of 10–20 counts in consecutive time intervals of 1-min duration for three or more preparations) of movements per minute per field (22 $\times$ 25 μ m) and the following rating system was used: minus sign, no movements; $\pm$, no movements, or <2 movements per min per field in some preparations; plus sign, >10 movements per min per field.

† The rating system for microfilament network formation is as follows: minus sign, no network around the edge of the bulk axoplasm; $\pm$, the absence of an extensive network with microfilaments only very close (10–20 μ m) to the bulk axoplasm; plus sign, extensive network extending over more than 100 μ m from bulk axoplasm.

‡ The rating system for appearance of free microtubules from bulk axoplasm is as follows: minus sign, no microtubules; $\pm$, very few microtubules compared with control; plus sign, same number of or more microtubules compared with control.

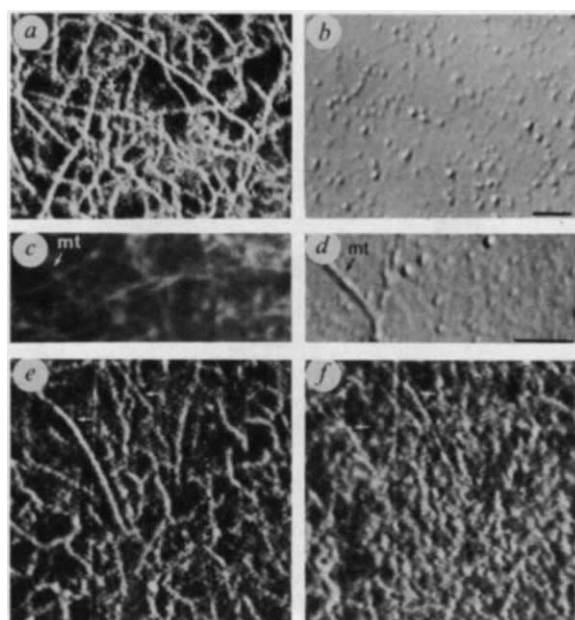
fluorescent filaments were not detected by Allen video-enhanced differential interference contrast (AVEC-DIC) microscopy (Fig. 1b) because they were thinner than the limit of detection (20 nm¹³). Therefore these were single actin filaments or bundles of not more than four or five. By AVEC-DIC microscopy, we observed organelles (Fig. 1b) and microtubules (Fig. 1d) that did not stain with rhodamine-phalloidin (Fig. 1c).

To examine network formation, the axoplasm was extruded directly into buffer containing rhodamine-phalloidin. The networks (Fig. 1e) were very similar to those seen after perfusion (Fig. 1a) but they were observed within 10–20 min instead of 30–40 min after extrusion, probably as a result of the stabilizing effect of phalloidin¹⁴. The network around the bulk axoplasm became denser with time (Fig. 1e,f), indicating that network

formation is a dynamic process. Extrusion of axoplasm into buffer containing 10 μ M cytochalasin B, a drug that prevents actin polymerization¹⁵, inhibited network formation not only in the absence but also in the presence of phalloidin, which binds to actin filaments and protects them against depolymerization (Table 1). These results support the view that networks form by actin assembly. Polymerization is possible in these preparations because the concentration of monomeric actin in axoplasm (~0.37 mg ml⁻¹; ref. 6) is above the critical concentration. Furthermore, depletion of both ATP and ADP by apyrase, and addition of EDTA (Table 1), treatments that destabilize or denature monomeric actin^{16,17}, blocked network formation. The omission of exogenous ATP from the buffer had no effect on network formation (Table 1). As squid axoplasm contains about

FIG. 1 Development of actin filament networks at the cover glass surface in extruded squid axoplasm preparations. a, Network visualized by rhodamine-phalloidin staining (0.4 μ M). b, Same field in AVEC-DIC microscopy shows attached particles but filaments are not detected. Comparison of the microfilament networks (a) with the pattern of stationary organelles in the same field (b) revealed that 90–95% of the organelles colocalized with the network if superimposed correctly, but only 60–70% colocalized when randomly superimposed. c, d, A microtubule (mt, arrowed) is visualized by AVEC-DIC microscopy but does not stain with rhodamine-phalloidin. e, f, Changes occur in the actin network with time; although some filaments (arrows) are still present, the network has become denser during the time interval of 12 min. a–d, Staining by perfusion of the preparation with rhodamine-phalloidin buffer. e–f, Network formed after extrusion into rhodamine-phalloidin buffer. Scale bars, 3 μ m.

METHODS. Squid (*Loligo pealei*) were obtained daily (Marine Resources, Woods Hole). Giant axons were dissected and extruded as previously described^{2,4} onto a microscope slide in 25 μ l axoplasmic dissociation buffer (100 mM potassium aspartate, 18 mM taurine, 10 mM betaine, 7.5 mM glycine, 4 mM HEPES, 2.0 mM MgCl₂ · 6H₂O, 1.4 mM CaCl₂, 1.4 mM EGTA, 2 mM ATP, pH 7.2). Slides were observed for 1–2 h after extrusion using fluorescence and AVEC-DIC microscopy according to Allen^{1,2,28}. For video imaging, either a Dage 67M (Dage-MTI, Michigan City) video camera with chalnicon tube or a Hamamatsu C2400 SIT camera (Hamamatsu Photonics) was used. The video signal was first subjected to analogue contrast enhancement^{2,29}, followed by digital image processing²⁹ with the Hamamatsu ARGUS 10 processor.



1 mM endogenous ATP¹⁸, hexokinase, an enzyme that hydrolyses ATP to ADP, was used to deplete the preparation of ATP. We found that the network formed with ADP and that ATP was not required (Table 1). Taken together, these findings suggest that network formation occurs mainly by actin polymerization and not by active gliding of filaments out of the axoplasm, for which ATP is required¹⁹.

In areas where individual microtubules were attached to the cover glass, characteristic bidirectional movement of organelles was observed along the microtubules. In the same areas, however, directed organelle movements also took place on the glass surface in places where no microtubules or tracks were visible by AVEC-DIC microscopy (Fig. 2*a, b*). The directed movement was linear, but characterized by abrupt changes in direction of up to 90° or more (Fig. 3*a–c*). The average length of linear excursions was $1.72 \pm 0.83 \mu\text{m}$ ($n = 225$). Some organelles moved over a total distance of 20 μm or more, while undergoing several turns as if moving on a polygonal pattern of invisible tracks (Fig. 3*a–c*). Some of the tracks were especially apparent because several organelles moved consecutively along them (Fig. 2*a*). All organelles on a given track moved in the same direction. Reversals of direction or two particles moving in opposite directions on the same track were never observed. As identical tracks were used for several minutes, the possibility that organelles were riding on gliding filaments is unlikely.

Organelles of various types and sizes moved rapidly along the invisible tracks, including small organelles with the low-contrast, characteristic of axoplasmic and synaptic vesicles² (Fig. 2*a*), larger round organelles of higher contrast, as described for prelysosomal organelles (Figs. 2*c, 3a, b*), and larger elongated structures with the morphology of mitochondria (Fig. 2*b*). The velocities varied slightly with organelle size (Fig. 3*e–g*). In fields where microtubule-dependent movement and the movement on invisible tracks coexisted, organelles sometimes switched from

a microtubule to invisible tracks (Fig. 2*c*). Therefore a given organelle can use both motility systems.

Organelle movements along tracks invisible with AVEC-DIC microscopy differ in several ways from organelle movements along microtubules. The speed of organelles moving on microtubules is faster by a factor of about 1.5, but in any single preparation it varies by only about 10% (refs 2, 3). Organelle movements on invisible tracks is similar to organelle movement in intact axoplasm where large organelles move more slowly and with a less continuous motion^{1,2}. Both the instantaneous velocities of individual particles and the average velocities of organelle populations were remarkably variable, having standard deviations between 40 and 60% (Fig. 3*d*). But unlike microtubule-based movement^{2,20}, the organelles moved only unidirectionally.

To determine whether the network of actin filaments constituted the tracks along which organelles moved, we observed the same fields after rhodamine-phalloidin staining by fluorescence and AVEC-DIC microscopy. The addition of the fluores-

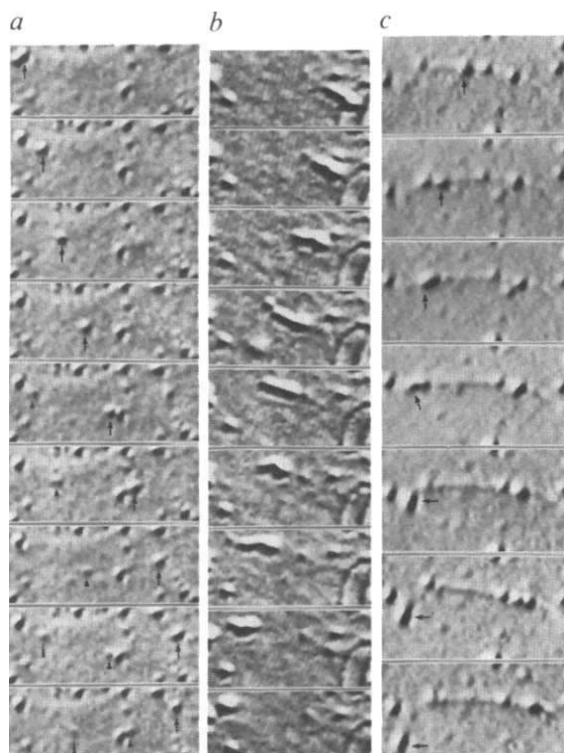


FIG. 2 Examples of different types of organelles moving along tracks invisible with AVEC-DIC microscopy. *a*, Three small-to-medium sized organelles (arrow and two arrowheads) move consecutively on the same track from left to right (intervals, 1 s). *b*, Mitochondrion-shaped organelle departing from a microtubule and moving over the cover glass surface (intervals, 1 s). *c*, Organelle (arrow) moves along a microtubule to the end and continues along an invisible track at the glass surface (intervals, 0.5 s). Scale bar, 2 μm .

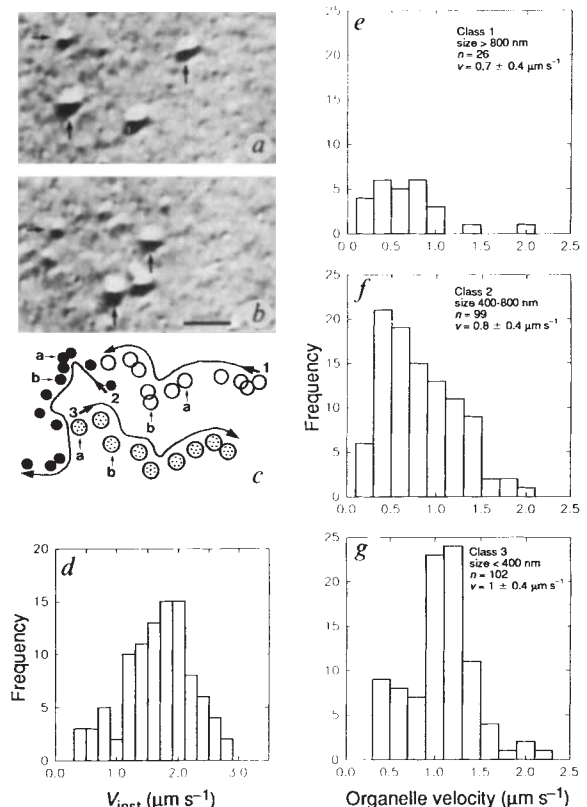


FIG. 3 Motion Analysis. *a*, Three large organelles moving on the coverglass surface. *b*, The same field 2 s later. *c*, Visualization of the tracks by depicting the particles' outlines at 1-s intervals. Lines with arrows in *c* show the direction only; the exact trajectory is obtained by connecting the centres of the particles. In *c*, the letters *a* and *b* mark the positions of the three particles in panels *a* and *b*. Scale bar, 2 μm . *d*, Distribution of the frequencies of instantaneous velocities (v_{inst}) attained during 12 s by one particle, that is, the velocities for the sampling intervals of 0.1 s. Mean and standard deviation are $1.7 \pm 0.6 \mu\text{m s}^{-1}$. *e–g*, Velocity histograms for the three size classes of organelles observed to move on invisible tracks. Sizes are apparent sizes because in AVEC-DIC microscopy objects as small as 25 nm are detectable, but actual size appears inflated by diffraction²⁹. Average velocities (v) were determined for episodes of straight motion between abrupt turns of direction. For each particle 3–6 episodes were examined. The 26 measurements in *e* were from 6 organelles, the 99 measurements in *f* from 34 organelles, and the 102 measurements in *g* from 42 organelles. METHODS. Motion was analysed by extracting the *x* and *y* coordinates of moving organelles from a series of video images either automatically (Hamamatsu XY-Tracker C1055) or by manually positioning a cursor and entering the coordinates. The following parameters were quantitatively evaluated using the software packages PARTI-MOVI and SPSS as described³: average velocity, instantaneous velocity, directionality and pauses.

cent probe did not impair organelle motility. The motile activity along invisible tracks occurred only in the presence of the network. Like the formation of networks, the movement of organelles along filaments invisible in AVEC-DIC microscopy began 30–40 min after extrusion in the absence of phalloidin and 10–20 min after extrusion when phalloidin was present. Cytochalasin B prevented not only network formation around the extruded axoplasm but also fully inhibited the directed motile activity (Table 1). Nocodazole, a drug that promotes microtubule disassembly²¹, had no effect on filamentous network formation and actin network-associated organelle movement, even at concentrations (100 μ M) that removed all microtubules in the vicinity of extruded axoplasm and therefore abolished microtubule-based organelle movements (Table 1). Thus, the new type of organelle movement depends on actin filaments and can occur in the absence of microtubules.

The effects of ATP and several ATPase inhibitors on actin-dependent organelle movement were determined (Table 1). The omission of exogenous ATP from the buffer had no effect on the motile activity, but the depletion of endogenous ATP by hexokinase resulted in a complete inhibition of organelle movement. These results provide evidence that actin-dependent organelle movement requires ATP and that an ATPase is involved in this process. The non-hydrolysable ATP analogue adenylyl imidodiphosphate (AMP-PNP), a potent inhibitor of kinesin-driven motility^{4,22}, did not inhibit the motile activity. Sodium orthovanadate, an inhibitor of dynein-driven motility^{23,24} also had no significant effect on organelle movement (Table 1). *N*-ethylmaleimide (1 mM), which inhibits myosin-driven motility²⁵, could not be used, because this treatment prevented the formation of the actin network. We conclude that neither dynein nor kinesin is responsible for organelle movement on actin filaments. These findings are compatible with a myosin-like motor, however, because according to kinetic studies on the actin-activated ATPase activity of myosin, AMP-PNP and vanadate are weak inhibitors of the myosin ATPase^{26,27}.

Our data show that actin filaments not only have a structural role, but also provide a second independent system for organelle transport in squid axoplasm. Both components of axonal transport appear to function in close proximity to each other because actin filaments are associated in axons with microtubule domains⁷, and because inhibitors that produce a rigor-like state of one of the two systems lead to an inhibition of all movements. Only outside the bulk axoplasm can the systems be separately inhibited. For example, AMP-PNP inhibits all organelle movement in axoplasm by freezing microtubule-based motility²⁸ but, as shown here, does not affect motility on free actin filaments. Conversely, gelsolin, DNase I, and dephosphorylated synapsin I inhibit all organelle movement in axoplasm^{10,11}, probably by blocking actin-based motility, whereas synapsin I and DNase I do not block the motility on free microtubules¹¹. Our observation that a given organelle can move along both microtubules and actin filaments (Fig. 2c) supports the hypothesis that the actin-based and microtubule-based motility systems are not only closely associated but also operate in close connection with each other to produce and regulate the movement of organelles. □

15. MacLean-Fletcher, S. & Pollard, T. D. *Cell* **20**, 329–341 (1980).
16. West, J. J., Nagy, B. & Gergely, J. *J. Biol. Chem.* **242**, 1140–1145 (1967).
17. West, J. J., Nagy, B. & Gergely, J. *Biochem. Biophys. Res. Commun.* **29**, 611–616 (1967).
18. Caldwell, P. C. *J. Physiol.* **152**, 561–590 (1960).
19. Kron, S. J. & Spudich, J. A. *Proc. natn. Acad. Sci. U.S.A.* **83**, 6272–6276 (1986).
20. Schnapp, B. J., Vale, R. D., Sheetz, M. P. & Reese, T. S. *Cell* **40**, 455–462 (1985).
21. Hoebeke, J., Van Nijen, G. & De Brabander, M. *Biochem. biophys. Res. Commun.* **69**, 319–324 (1976).
22. Cohn, S. A., Ingold, A. L. & Scholey, J. M. *Nature* **328**, 160–163 (1987).
23. Gibbons, I. R. et al. *Proc. natn. Acad. Sci. U.S.A.* **75**, 2220–2224 (1978).
24. Paschal, B. M. & Vallee, R. B. *Nature* **330**, 181–183 (1987).
25. Sheetz, M. P. & Spudich, J. A. *Nature* **303**, 31–35 (1983).
26. Greene, L. E., Sellers, J. R., Eisenberg, E. & Adelstein, R. S. *Biochemistry* **22**, 530–535 (1983).
27. Smith, S. J. & Eisenberg, E. *Eur. J. Biochem.* **193**, 69–73 (1990).
28. Brady, S. T., Lasek, R. J. & Allen, R. D. *Cell Motil.* **5**, 81–101 (1985).
29. Weiss, D. G., Maile, W. & Wick, R. A. in *Light Microscopy in Biology—A Practical Approach* (ed. Lacey, A. J.) 221–278 (IRL, Oxford, 1989).

ACKNOWLEDGEMENTS. We thank S. Authier, K. Weissbacher and W. Maile for help with image analysis, motion analysis and image processing, G. Manley and D. Murphy for support and for critically reading the manuscript, and D. Rivera and D. Cohen for assistance with experiments at Woods Hole. This work was supported by a fellowship to S.A.K. from the A.-v. Humboldt Foundation, grants from the NSF (G.M.L.) and DFG (D.G.W.), and by an R.D. Allen Fellowship of the MBL (S.A.K.).

A eukaryotic DNA glycosylase/lyase recognizing ultraviolet light-induced pyrimidine dimers

Krista K. Hamilton, Peter M. H. Kim
& Paul W. Doetsch*

Department of Biochemistry, Emory University School of Medicine,
Rollins Research Center, Atlanta, Georgia 30322, USA

CYCLOBUTANE pyrimidine dimers (CPDs) are the predominant product of photodamage in DNA after exposure of cells to ultraviolet light^{1,2} and are cytotoxic, mutagenic and carcinogenic in a variety of cellular and animal systems^{3–5}. In prokaryotes, enzymes and protein complexes have been characterized that remove or reverse CPDs in DNA^{6–8}. *Micrococcus luteus* and T4 phage-infected *Escherichia coli* contain a specific *N*-glycosylase/apurinic-apyrimidinic lyase that catalyses a two-step DNA incision process at sites of CPDs, thus initiating base excision repair of these lesions^{7,9–12}. It is well established that CPDs are recognized and removed from eukaryotic DNA by excision repair processes but very little information exists concerning the nature of the proteins involved in CPD recognition and DNA incision events^{7,12,13}. We report here that an enzyme functionally similar to the prokaryotic *N*-glycosylase/apurinic-apyrimidinic lyases exists in *Saccharomyces cerevisiae*. To our knowledge, this is the first time such an activity has been found in a eukaryote and is also the first example of an organism having both direct reversal and base excision repair pathways for the removal of CPDs from DNA.

We used a highly specific DNA cleavage assay based on DNA sequencing methodologies to characterize DNA scission events and have previously identified a yeast enzyme, redoxendonuclease, that recognizes and excises pyrimidine hydrate-type photoproducts induced by high doses of ultraviolet radiation and other types of oxidative DNA damage¹⁴. During a large-scale purification of redoxendonuclease from bakers' yeast, we detected an additional distinct activity after phosphocellulose chromatography that seemed to incise ultraviolet-irradiated DNA only at positions of adjacent pyrimidines, suggesting that cleavage at CPDs might be taking place (not shown). Fractions containing the putative CPD-incising activity, which we have termed yeast pyrimidine dimer endonuclease, were subjected to gel filtration chromatography to resolve this pyrimidine dimer endonuclease from yeast redoxendonuclease (Fig. 1).

* To whom correspondence should be addressed.

Received 10 January; accepted 9 March 1992.

1. Allen, R. D., Metzuzals, J., Tasaki, I., Brady, S. T. & Gilbert, S. P. *Science* **218**, 1127–1128 (1982).
2. Allen, R. D. et al. *J. Cell Biol.* **100**, 1736–1752 (1985).
3. Weiss, D. G., Langford, G. M., Seitz-Tutter, D. & Keller, F. *Cell Motil. Cytoskel.* **10**, 285–295 (1988).
4. Vale, R. D., Reese, T. S. & Sheetz, M. P. *Cell* **42**, 39–50 (1985).
5. Gilbert, S. P. & Sloboda, R. D. *J. Cell Biol.* **109**, 2379–2394 (1989).
6. Morris, J. R. & Lasek, R. J. *J. Cell Biol.* **98**, 2064–2076 (1984).
7. Fath, K. R. & Lasek, R. J. *J. Cell Biol.* **107**, 613–621 (1988).
8. Goldberg, D. J., Harris, D. A., Lubit, B. W. & Schwartz, J. H. *Proc. natn. Acad. Sci. U.S.A.* **77**, 7448–7452 (1980).
9. Isenberg, G., Schubert, P. & Kreutzberg, G. W. *Brain Res.* **194**, 588–593 (1980).
10. Brady, S. T., Lasek, R. J., Allen, R. D., Yin, H. L. & Stossel, T. P. *Nature* **310**, 56–58 (1984).
11. McGuinness, T. L. et al. *J. Neurosci.* **9**, 4138–4149 (1989).
12. Wulf, E., Debohen, A., Bautz, F. A., Faulstich, H. & Wieland, T. *Proc. natn. Acad. Sci. U.S.A.* **76**, 4498–4502 (1979).
13. Weiss, D. G. *J. Cell Sci.* **55**, 1–15 (1986).
14. Estes, J. E., Selden, L. A. & Gershman, L. C. *Biochemistry* **20**, 708–712 (1981).

To characterize the yeast pyrimidine dimer endonuclease we used T4 endonuclease V (T4 endo V) because of its known CPD-incising activity, and *E. coli* DNA photolyase which, in the presence of visible light, reverses CPDs to their normal constitutive pyrimidines¹⁵. Ultraviolet-damaged 3' end-labelled DNA fragments of defined sequence were reacted with DNA photolyase or left untreated and both types of substrates were then incubated with either the yeast endonuclease or T4 endo V and the reaction products analysed on DNA-sequencing gels (Fig. 2). As expected, T4 endo V incised such substrates at sites of CPDs to produce DNA scission products (lane 4) with electrophoretic mobilities slightly above the DNA-sequencing chemical cleavage fragments corresponding to the 5' pyrimidine of each CPD (lanes CT and C). The same substrate incubated with the yeast endonuclease produced a nearly identical pattern

of DNA cleavage products (lane 1), indicating that recognition and incision at CPDs might be similar to T4 endo V. The positions of other non-CPD ultraviolet photoproducts (pyrimidine photohydrates (lane 7) and (6-4) pyrimidine-pyrimidone photoproducts (lane 9)) were also mapped using lesion-specific DNA cleavage techniques¹⁶ and resulted in completely different patterns of DNA scission products compared with those produced by the yeast endonuclease and T4 endo V. Reversal of CPDs contained in these substrates with *E. coli* DNA photolyase prevented cleavage of DNA by both endonucleases, confirming that the yeast pyrimidine dimer endonuclease recognizes and cleaves DNA at sites with CPDs. The size of the yeast enzyme is in the M_r range 16–20K and it does not seem to require divalent cations for activity. In these respects, it is similar to both T4 endo V and the *M. luteus* ultraviolet endonuclease^{11,17}.

T4 endo V processes CPDs through an *N*-glycosylase/apurinic-apyrimidinic (AP) lyase mechanism^{10,18–21} (Fig. 4a). DNA strand cleavage experiments were done using depurinated, 5' end-labelled DNA substrates to determine whether or not the yeast endonuclease had AP lyase properties similar to T4 endo V (Fig. 3). Depurinated DNA substrates were incubated with the yeast enzyme (lane 2), T4 endo V (lane 3), *E. coli* endonuclease IV (an AP endonuclease; lane 4) or were

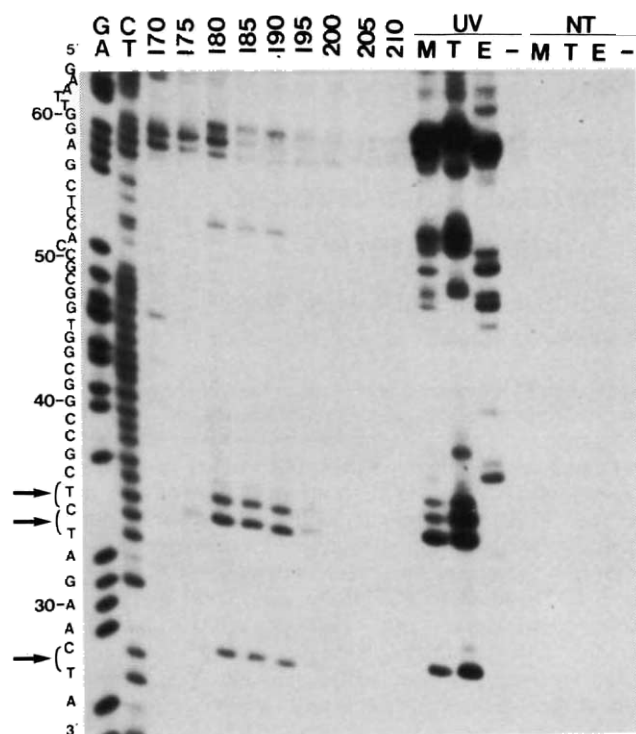


FIG. 1 Detection of yeast pyrimidine dimer endonuclease (YPDE) activity during Sephacryl S-200 chromatography of yeast redoxendonuclease. An ultraviolet-irradiated (10 kJ m^{-2}) 3' end-labelled DNA fragment (fragment A) (179 base pairs (bp)) was incubated with 25 μl of the indicated S-200 column fractions and the reaction products were analysed on a denaturing 20% polyacrylamide DNA-sequencing gel¹⁴. Ultraviolet-damaged (UV lanes) and undamaged (NT lanes) substrates were also incubated with *M. luteus* ultraviolet endonuclease (M lanes, 10 μg ; Applied Genetics) or T4 endo V (T lanes, 0.6 ng) or *E. coli* endonuclease III (E lanes, 60 ng), or left untreated (—) as previously described^{10,14,31}. (G+A) and (C+T) DNA-sequencing reactions³² are shown in the far left lanes. Arrows indicate YPDE-mediated cleavage at sites of CPDs. Most of the radioactivity in each lane corresponded to uncleaved, full-length DNA fragments and was located at the top of the gel (not shown). Base numbering is from the 3' end-labelled terminus. METHODS. The purification of YPDE will be described in detail elsewhere. Crude yeast extracts (fraction I) were prepared from 340 g dried bakers' yeast (Red Star, Universal Foods) by mechanical disruption and subjected to DEAE-cellulose (fraction II) and phosphocellulose chromatography (fraction III) as previously described for the purification of yeast redoxendonuclease¹⁴. YPDE copurified with yeast redoxendonuclease through fraction III, which was then size-fractionated by Sephacryl S-200 chromatography (2 ml fractions) as described¹⁴. The peak YPDE fractions (185–190) were pooled, concentrated and constituted fraction IV (total protein, 3 mg) and used in subsequent experiments. Fraction IV produced five major bands on a silver-stained SDS-polyacrylamide gel. The DNA repair enzyme substrate (fragment A) was a 179-bp *EcoRI*–*PvuII* restriction fragment generated from pBluescript and was 3' ³²P end-labelled as described¹⁴. Ultraviolet irradiation (254 nm) of substrates has been described¹⁴.

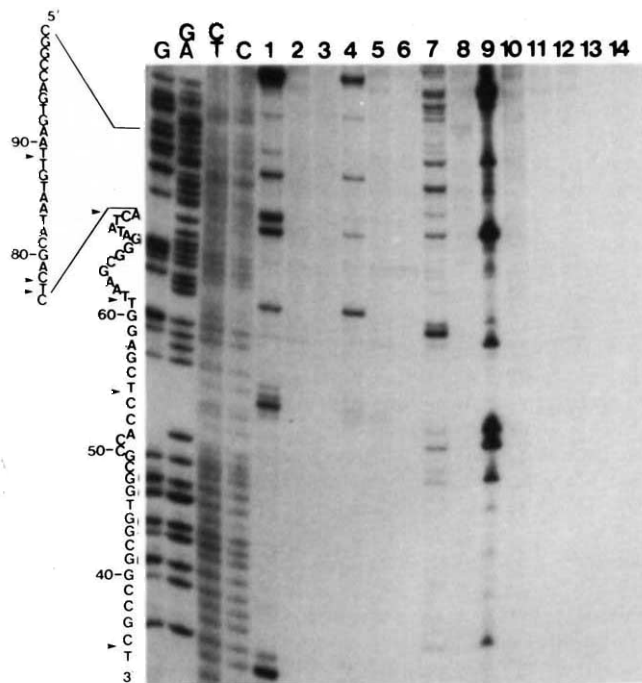


FIG. 2 Yeast pyrimidine dimer endonuclease (YPDE) cleavage of ultraviolet-irradiated DNA at sites of CPDs and comparison to T4 endo V. The positions of non-CPD ultraviolet photoproducts were determined by treatment of the ultraviolet-damaged, end-labelled substrates with hot alkali or *E. coli* endonuclease III to obtain DNA cleavage events at sites of (6-4) pyrimidine-pyrimidone photoproducts and monobasic pyrimidine photohydrates, respectively¹⁶. Ultraviolet-irradiated (2 kJ m^{-2} ; lanes 1, 2, 4, 5, 7, 9, 11, 12) and undamaged (lanes 3, 6, 8, 10, 13, 14) 3' end-labelled DNA fragment A was treated with *E. coli* DNA photolyase (0.1 μg) and photoreactivating light to reverse CPDs as described³³ (lanes 2, 5, 12, 13) and/or incubated with YPDE (fraction IV, 5 μg ; lanes 1, 2, 3); T4 endo V (lanes 4, 5, 6); endonuclease III (lanes 7, 8); hot alkali (lanes 9, 10); or no treatments (lanes 11, 14). Base numbering starts from the 3'-labelled terminus. Reaction products were analysed on DNA sequencing gels and DNA sequencing reactions were run alongside in the far left lanes as in Fig. 1. Small arrows indicate cleavages occurring at sites of CPDs.

METHODS. T4 endo V and endonuclease III incubations were as described in the legend to Fig. 1. Hot alkali (1M piperidine, 90 °C, 30 min) cleavage of ultraviolet-damaged DNA at sites of (6-4) photoproducts was done as described¹⁶. Ultraviolet irradiation of DNA fragments was as described in the legend to Fig. 1.

treated with hot alkali (lane 1). All four procedures resulted in DNA cleavage at AP sites. The differences in the electrophoretic mobilities of the DNA scission products generated by these treatments can be accounted for by differences in the nature of the 3' termini produced after DNA strand cleavage. The yeast and T4 endonucleases also generated identical cleavage products when depurinated 3' end-labelled substrates were used (not shown). The results of these experiments are consistent with both enzymes cleaving DNA at AP sites through a β -elimination reaction.

The ability of the yeast pyrimidine dimer endonuclease to recognize and cleave CPDs and AP sites in a manner similar to T4 endo V suggested that the yeast enzyme might also cleave the *N*-glycosidic bond of the 5' pyrimidine of CPDs as does T4 endo V. A strategy was devised to reveal any such *N*-glycosylase activity assuming different electrophoretic mobilities for a DNA cleavage fragment with a CPD attached to the 5' terminus and the same fragment without a CPD (Fig. 4a). A parallel set of experiments was done with ultraviolet-damaged, 3' end-labelled DNA substrates incubated with each nuclease and the resulting DNA cleavage products (with and without subsequent DNA photolyase treatments) were analysed. DNA photolyase treatment of substrates cleaved by either the yeast or the T4 endonuclease (Fig. 4b and c, respectively) produced DNA scission products that migrated slightly ahead of their untreated counterparts to positions equivalent to those of the chemical cleavage DNA sequencing fragments for the 5' pyrimidine of each CPD (compare plus and minus lanes YPDE and T4). For DNA

substrates incubated with the yeast enzyme, such a result would only be possible if *N*-glycosidic cleavage of 5' pyrimidines of CPDs were taking place, and we conclude that both the yeast and T4 endonucleases process CPDs by very similar if not identical mechanisms.

S. cerevisiae has been used extensively as a model eukaryotic system for studying the genetics of DNA repair. About 90 genes have been identified that are involved in the yeast cellular response to DNA damage²². These genes have been grouped into three epistasis groups (*RAD3*, *RAD6* and *RAD52*) representing different responses to ultraviolet and/or ionizing radiation-induced DNA damage (reviewed in ref. 22). So far only the products of the *RAD3*, *RAD6* and *RAD9* genes have been characterized, and these correspond to a DNA helicase²³, a ubiquitin-conjugating enzyme²⁴ and a factor that controls cell-cycle progression²⁵, respectively. Whether any of the known *rad* mutants lacks yeast pyrimidine dimer endonuclease activity is not yet known, but the identification of this enzyme in yeast suggests that a similar enzyme might exist in other eukaryotes. Previous investigations may have failed to identify such enzymes because of the methods used for detection of CPD-specific *N*-glycosylase/AP lyase activities or because of enzyme lability^{12,13,26,27}.

The results of these studies indicate that *S. cerevisiae* is able to repair CPDs by direct reversal (DNA photolyase-mediated) and base-excision repair (mediated by the pyrimidine dimer endonuclease) pathways. To our knowledge, no other organism has been shown to possess both of these pathways for the repair

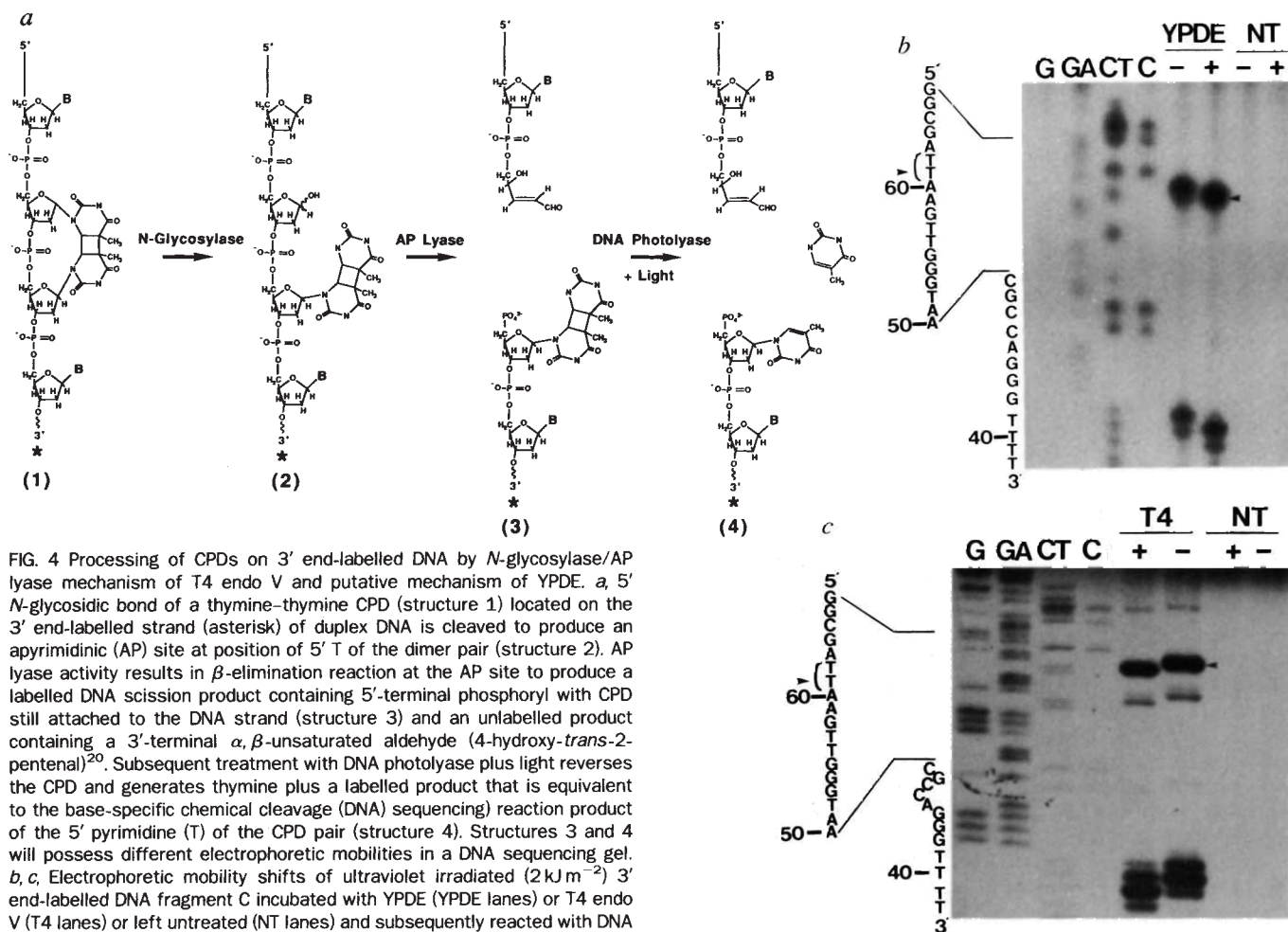


FIG. 4 Processing of CPDs on 3' end-labelled DNA by *N*-glycosylase/AP lyase mechanism of T4 endo V and putative mechanism of YPDE. a, 5' *N*-glycosidic bond of a thymine-thymine CPD (structure 1) located on the 3' end-labelled strand (asterisk) of duplex DNA is cleaved to produce an apyrimidinic (AP) site at position of 5' T of the dimer pair (structure 2). AP lyase activity results in β -elimination reaction at the AP site to produce a labelled DNA scission product containing 5'-terminal phosphoryl with CPD still attached to the DNA strand (structure 3) and an unlabelled product containing a 3'-terminal α,β -unsaturated aldehyde (4-hydroxy-*trans*-2-pentenal)²⁰. Subsequent treatment with DNA photolyase plus light reverses the CPD and generates thymine plus a labelled product that is equivalent to the base-specific chemical cleavage (DNA sequencing) reaction product of the 5' pyrimidine (T) of the CPD pair (structure 4). Structures 3 and 4 will possess different electrophoretic mobilities in a DNA sequencing gel. b, c, Electrophoretic mobility shifts of ultraviolet irradiated (2 kJ m⁻²) 3' end-labelled DNA fragment C incubated with YPDE (YPDE lanes) or T4 endo V (T4 lanes) or left untreated (NT lanes) and subsequently reacted with DNA photolyase plus light (+ lanes). Arrow indicates DNA cleavage event occurring at CPD site T₆₁-T₆₂. Reaction products were electrophoresed on DNA sequencing gels and DNA sequencing reactions were included as in Fig. 1. Base numbering is from the 3' end-labelled terminus.

METHODS. YPDE, T4 endo V and DNA photolyase (plus light) incubations are

described in the legend to Fig. 2 and ultraviolet irradiation of DNA fragments in the legend to Fig. 1. DNA repair enzyme substrate fragment C was a 91-bp *EcoRI*-*PvuII* restriction fragment generated from pUC19 and 3' ³²P end-labelled as described¹⁴.

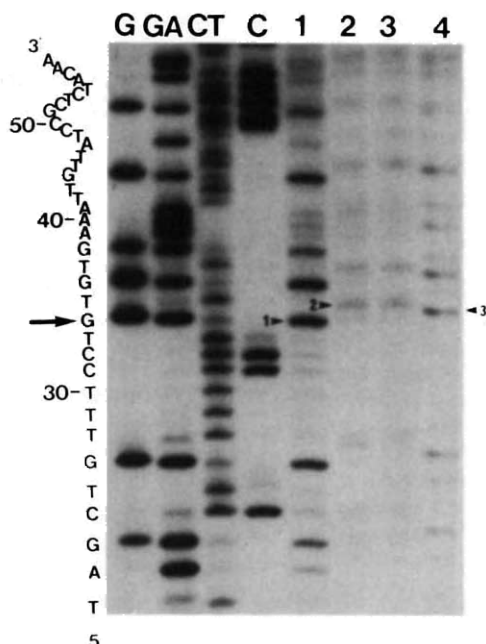


FIG. 3 Apurinic/aprimidinic (AP) lyase activity associated with yeast pyrimidine dimer endonuclease (YPDE). Depurinated, 5' end-labelled DNA fragment B was treated with hot alkali (lane 1), YPDE (lane 2), T4 endo V (lane 3) or *E. coli* endonuclease IV (10 ng, lane 4). Large arrow indicates DNA cleavage event at AP site G₃₄ by hot alkali (small arrow 1) producing a 3' phosphoryl terminus, or YPDE and T4 endo V (small arrow 2) producing a 3' modified deoxyribose terminus, or endonuclease IV (small arrow 3) producing a 3' hydroxyl terminus. YPDE and T4 endo V cleaved depurinated 5' end-labelled DNA at AP sites and produced DNA scission products with identical electrophoretic mobilities, suggesting that the nature of the 3' termini are the same for both enzymes and correspond to a modified deoxyribose. By contrast, hot alkali cleavage at AP sites produced DNA cleavage products containing 3'-phosphoryl groups¹⁶ which migrated ahead of the YPDE and T4 endo V-generated cleavage products and aligned with the corresponding base-specific chemical cleavage products in the DNA sequencing lanes (lanes G and G+A). *E. coli* endonuclease IV, an enzyme that cleaves duplex DNA 5' to AP sites generating 3'-hydroxyl and 5' deoxyribose-5'-phosphate termini³⁴ was used as an additional AP endonuclease control and generated DNA scission products with electrophoretic mobilities intermediate to those produced by YPDE (or T4 endo V) and hot alkali treatment. Reaction products were electrophoresed on DNA sequencing gels and DNA sequencing reactions were included in the far left lanes as in Fig. 1. Base numbering is from the 5'-labelled terminus.

METHODS. End-labelled DNA repair enzyme substrates were depurinated as described³⁵. Indicated incubations were done as previously described for hot alkali and endo III (ref. 14), and endonuclease IV (ref. 36). YPDE and T4 endo V incubations were as described in the legend to Fig. 2. DNA repair enzyme substrate fragment B was a 179-bp *HindIII*-*PvuII* restriction fragment generated from pUC19 and 5' ³²P end-labelled as described¹⁴.

of CPDs. There is also evidence from studies with *RAD3* epistasis group mutants that indicates yeast is able to remove CPDs from DNA through a nucleotide excision repair pathway²². The finding that a number of *RAD3* epistasis group mutants are defective in incision of ultraviolet-damaged DNA^{22,28-30} suggests that the interaction of several proteins is required for CPD removal in yeast, whether it is performed by the base excision repair (yeast pyrimidine dimer endonuclease-mediated) or nucleotide excision repair pathways. Some of the *RAD3* genes also function in other DNA processing events (for example, recombination) and certain *RAD* gene products may even mediate critical functions for elements of both excision repair pathways. Hence *in vivo* the pyrimidine dimer endonuclease may be necessary but not sufficient for efficient excision of CPDs via the base-excision repair pathway. As mutants of the enzyme and more information on the function and relationships of the *RAD* gene products becomes available, the relative contributions of these two pathways in the excision repair of CPDs should become clearer. □

Received 23 December 1991; accepted 28 February 1992.

- Patrick, M. H. & Gray, D. *Photochem. Photobiol.* **24**, 507-513 (1976).
- Patrick, M. H. *Photochem. Photobiol.* **25**, 357-372 (1977).
- Harm, W. *The Biological Effects of Ultraviolet Radiation* (Cambridge Univ. Press, 1980).
- Cochill, T. P. *Photochem. Photobiol.* **46**, 1043-1050 (1987).
- Hart, R. W. & Setlow, R. B. *Molecular Mechanisms of DNA Repair Part B* 719-724 (Plenum, New York, 1975).
- Sancar, A. & Sancar, G. B. *A. Rev. Biochem.* **57**, 29-67 (1988).
- Friedberg, E. C. *DNA Repair* (Freeman, New York, 1985).
- Sancar, G. B. *Mutat. Res.* **236**, 147-160 (1990).
- Dodson, M. L. & Lloyd, R. S. *Mutat. Res.* **218**, 49-65 (1990).
- Haseltine, W. A. *et al. Nature* **285**, 634-641 (1980).
- Grafstrom, R. H., Park, L. & Grossman, L. *J. Biol. Chem.* **257**, 13465-13474 (1982).
- Radany, E. H. & Friedberg, E. C. *J. Virol.* **41**, 88-96 (1982).
- Friedberg, E. C. *et al. J. Supramolec. Struct. Cell. Biochem.* **16**, 91-103 (1981).
- Gossett, J., Lee, K., Cunningham, R. P. & Doetsch, P. W. *Biochemistry* **27**, 2629-2634 (1988).
- Sancar, A., Smith, F. W. & Sancar, G. B. *J. Biol. Chem.* **259**, 6028-6032 (1984).
- Doetsch, P. W., Helland, D. E. & Haseltine, W. A. *Biochemistry* **25**, 2212-2220 (1986).
- Friedberg, E. C., Ganesan, A. K. & Seawall, P. C. *Meth. Enzym.* **65**, 191-201 (1980).
- Radany, E. H. & Friedberg, E. C. *Nature* **286**, 182-185 (1980).
- Seawall, P. C., Smith, C. A. & Ganesan, A. K. *J. Virol.* **35**, 790-792.
- Mazumder, A. *et al. Am. Chem. Soc.* **111**, 8029-8030 (1989).
- Gordon, L. K. & Haseltine, W. A. *J. Biol. Chem.* **255**, 12047-12050 (1980).
- Friedberg, E. C. *Microbiol. Rev.* **52**, 70-102 (1988).
- Sung, P., Prakash, L., Matson, S. W. & Prakash, S. *Proc. natn. Acad. Sci. U.S.A.* **84**, 8951-8955 (1987).
- Jentsch, S., McGrath, J. P. & Varshavsky, A. *Nature* **329**, 131-134 (1987).
- Weinert, T. A. & Hartwell, L. H. *Science* **241**, 317-322 (1988).
- Bekker, M. L., Kaboev, O. K., Akhmedov, A. T. & Luchkina, L. A. *J. Bact.* **142**, 322-324 (1980).
- Waldstein, E. A., Peller, S. & Setlow, R. B. *Proc. natn. Acad. Sci. U.S.A.* **76**, 3746-3750 (1979).
- Prakash, L. *Mutat. Res.* **45**, 13-20 (1977).
- Reynolds, R. J. & Friedberg, E. C. *J. Bact.* **146**, 692-704 (1981).
- Bankman, M., Prakash, L. & Prakash, S. *Nature* **355**, 555-558 (1992).
- Dowd, D. R. & Lloyd, R. S. *J. Biol. Chem.* **265**, 3424-3431 (1990).
- Maxam, A. M. & Gilbert, W. *Meth. Enzym.* **65**, 499-560 (1980).
- Payne, G., Heelis, P. F., Rohrs, B. R. & Sancar, A. *Biochemistry* **26**, 7121-7127 (1987).
- Baily, V. & Verly, W. G. *Biochem. J.* **259**, 761-768 (1989).
- Hasegawa, S. L. *et al. Nucleic Acids Res.* **19**, 4915-4920 (1991).
- Levin, J. D. & Dimple, B. J. *Biol. Chem.* **263**, 8066-8071 (1988).

ACKNOWLEDGEMENTS. We thank A. Sancar for *E. coli* DNA photolyase, R. S. Lloyd for T4 endo V, R. P. Cunningham for *E. coli* endonuclease III and B. Dimple for *E. coli* endonuclease IV. This work was supported by the NIH. P.W.D. was supported by a Research Career Development Award from the NCI and K.K.H. was supported by an NIH Training Grant.

GUIDE TO AUTHORS

PLEASE follow these guidelines so that your manuscript may be handled expeditiously.

Nature is an international journal covering all the sciences. Contributions should be written in English, which is the language of the journal and for which English is a second language, and write clearly and simply, avoiding unnecessary technical terminology. Space in the journal is limited, making competition for publication severe. Brevity is highly valued. One printed page of *Nature*, without illustrations or tables, contains about 1,300 words.

Manuscripts are selected for publication according to editorial assessment of their suitability and reports from independent referees. They can be sent to London or Washington and should be addressed to the Editor. Manuscripts may be dealt with in either office, without interruption of the text, and where necessary be sent between offices by overnight courier. All manuscripts are acknowledged on receipt but fewer than half are sent for review. Those that are not reviewed are returned as rapidly as possible so that they may be submitted elsewhere without delay. Contributors may suggest reviewers; limited requests for the exclusion of specific reviewers are usually heeded. Manuscripts are usually sent to two or three reviewers, who are chosen for their expertise rather than their geographical location. Manuscripts accepted for publication are typeset from the London office.

Nature requests authors to deposit sequence and crystallographic data in the databases that exist for this purpose, and to mention availability of these data.

Once a manuscript is accepted for publication, contributors will receive galley proofs in about 4 weeks. *Nature's* staff will edit manuscripts with a view to brevity and clarity, so contributors should check their proofs carefully. Manuscripts are generally published 3-4 weeks after receipt of corrected proofs. *Nature* does not exact page charges. Contributors receive a reprint order form with their proofs; reprint orders are processed after the manuscript is published and payment received.

Categories of paper

Review articles survey recent developments in a field. Most are commissioned, but suggestions are welcome in the form of a one-page synopsis addressed to the Reviews Coordinator. Length is negotiable in advance.

Articles are research reports whose conclusions are of general interest and which are sufficiently rounded to be a substantial advance in understanding. They should not have more than 3,000 words of text (not including figure legends) or more than six display items (figures and tables) and should not occupy more than five pages of *Nature*.

Articles start with a heading of 50-80 words written to advertise their content in general terms, to which editors will pay particular attention. The heading does not usually contain numbers, abbreviations or measurements. The introduction to the study is contained in the first two or three paragraphs of the article, which also briefly summarizes its results and implications. Articles have fewer than 30 references and may contain a few subheadings of two or three words. **Letters** are short reports of outstanding novel findings whose implications are general and important enough to be of interest to those outside the field. Letters should have 1,000 or fewer words of text and four or fewer display items. The first paragraph describes, in not more than 50 words, the results and conclusions of the study, the background, rationale and chief conclusions of the study, for the particular benefit of non-specialist readers. Letters do not have subheadings and contain fewer than 30 references.

Commentary articles deal with issues in, or arising from, research that are also of interest to readers outside research. Some are commissioned but suggestions can be made to the commentary editor in the form of a one-page synopsis. Commentaries are normally between one and four pages of *Nature*.

News and Views articles inform non-specialist readers about new scientific advances, sometimes in the form of a conference report. Most are commissioned but proposals can be made in advance to the

News and Views editor.

Scientific Correspondence is for discussion of topical scientific matters, including those published in *Nature*, and for miscellaneous contributions. Priority is given to letters of fewer than 500 words and 5 references.

Preparation of manuscripts

All manuscripts should be typed, double-spaced, on one side of the paper only. An original and three copies are required, each accompanied by artwork. If photographs are included, four sets of originals are required: for line drawings, one set of originals and three good-quality photocopies are acceptable. Reference lists, figure legends and tables should all be on separate sheets, all of which should be double-spaced and numbered. Relevant manuscripts in press or submitted for publication elsewhere should be included with each copy of a submitted manuscript, and clearly marked as such. Revised and resubmitted manuscripts should also be clearly marked as such and labelled with their manuscript numbers.

Titles say what the paper is about with the minimum of technical terminology and in fewer than 80 characters in the case of Articles and Letters. Active verbs, numerical values, abbreviations and punctuation are to be avoided. Titles should contain one or two key words for indexing purposes.

Abstract should be marked individually and clearly with the author's name and, when known, the manuscript number. Ideally, no figure should be larger than 28 by 22 cm. Figures with several parts are to be avoided and are permitted only if the parts are closely related, either experimentally or logically. Unlettered originals of photographs should be provided. Original artwork is returned when a manuscript cannot be published.

Protein/nucleotide sequences should ideally be in the three-letter and not the single-letter code for amino acids. One column width of *Nature* can accommodate 20 amino acids or 60 base pairs. Numbering of sequences should be in the left-hand margin only, with a single space between rows.

Suggestions for cover illustrations, with captions and labelled with the manuscript number, are welcome.

Colour artwork. A charge of £300 per page is made as a contribution towards the cost of reproducing colour figures. Inability to pay these costs will not prevent the publication of essential colour figures if the circumstances are explained. Proofs of colour artwork may be sent to contributors under separate cover from their galley proofs.

Figure legends should not exceed 300 words and ideally should be shorter. The figure is described first, then, briefly, the method. Reference to a method published elsewhere is preferable to a full description. Methods are not described in the text.

References are numbered sequentially as they appear in the text, followed by those in tables and finally by those in figure legends. Only papers published or in the press are numbered and included in the reference list. All other forms of reference, including unrefereed abstracts, should be cited in the text as a personal communication, manuscript submitted or in preparation. Text is not included in reference lists. References are abbreviated according to the *World List of Scientific Periodicals* (Butterworths, London, 1963-65). The first and last page numbers are included: reference to books should include publisher, place and date.

Abbreviations, symbols, units and Greek letters should be identified the first time they are used. Acronyms should be avoided whenever possible and, if used, defined. Footnotes are not used except for changed addresses.

Acknowledgements are brief: grant and contribution numbers are not allowed.

Submission. Manuscripts can be sent to the Editor at 4 Little Essex Street, London WC2R 3LF, UK or at 1137 National Press Building, Washington, DC 20045, USA. Manuscripts or proofs sent by air courier to London should be declared as 'manuscripts' and valued \$5 to prevent the imposition of import duty and value-added tax (VAT). □